NASA
TT
F-268
c.1

P. N. Tverskoi

LOAN COPY: RETU
AFWL (WLIL-2
KIRTLAND AFB, N

TECH LIBRARY KAFB, NM
0068600

PHYSICS OF THE ATMOSPHERE

A Course in Meteorology

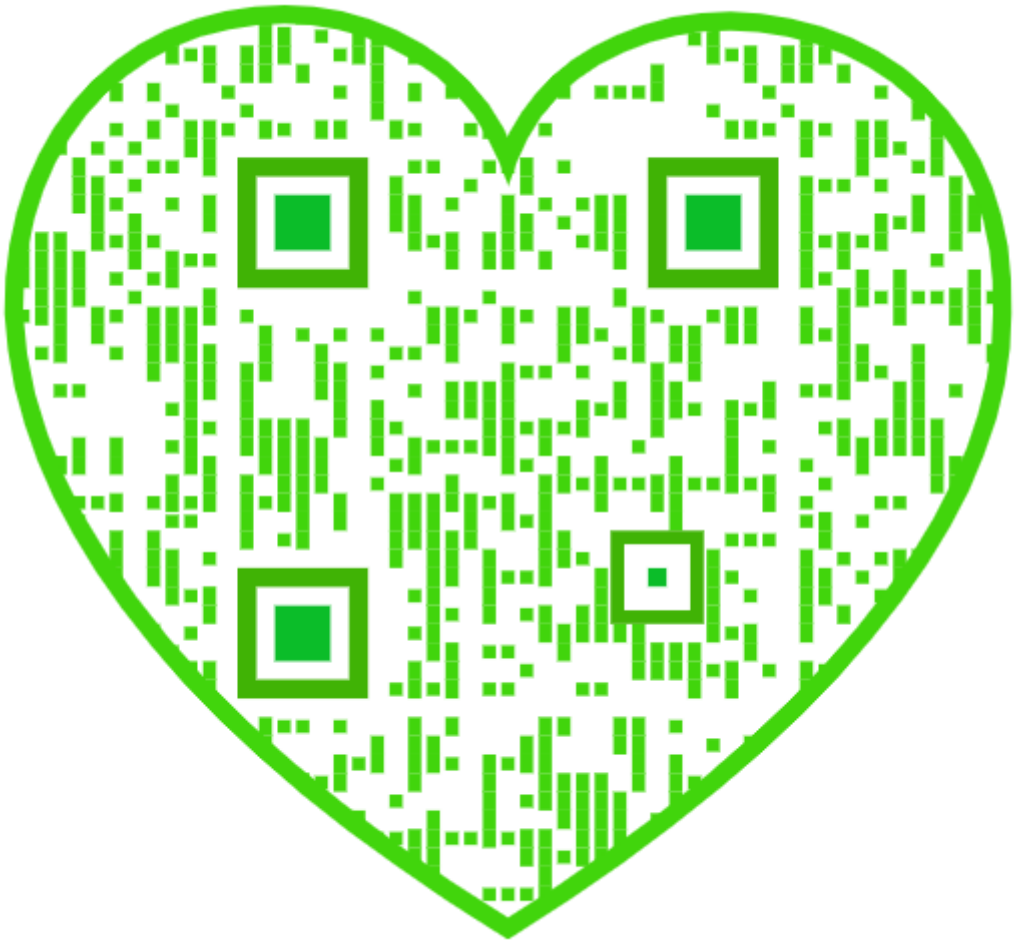
RECEIVED
4 MAY 1966
Technical Library
AF Weapons Laboratory

Translated from Russian

Published for the National Aeronautics and Space Administration, U.S.A.

and the National Science Foundation, Washington, D.C.

by the Israel Program for Scientific Translations



TECH LIBRARY KAFB, NM
0068600

P.N. TVERSKOI

PHYSICS OF THE ATMOSPHERE

A Course in Meteorology

(KURS METEOROLOGII)
(Fizika atmosfery)

E.S. Selezneva, Editor

Approved by the Ministry of
Higher and Secondary Specialized
Education of the USSR as a textbook
for hydrometeorological institutes
and universities

GIMIZ
Gidrometeorologicheskoe Izdatel'stvo
Leningrad 1962

Translated from Russian

Israel Program for Scientific Translations
Jerusalem 1965

NASA TT F-288
TT 65-50114

Published Pursuant to an Agreement with
THE NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, U.S.A.
and
THE NATIONAL SCIENCE FOUNDATION, WASHINGTON, D.C.

Copyright © 1965
Israel Program for Scientific Translations Ltd.
IPST Cat. No. 1398

Translated by A. Sen and R. N. Sen

Printed in Jerusalem by S. Monson
Binding: K. Wiener

Price: $8.76

Available from the
U.S. DEPARTMENT OF COMMERCE
Clearinghouse for Federal Scientific and Technical Information
Springfield, Va. 22151

TABLE OF CONTENTS

PREFACE xi

Chapter 1. INTRODUCTION 1
§ 1. The subject-matter of meteorology 1
§ 2. Relationship between meteorology and other sciences; subdivisions of meteorology . . 2
§ 3. Practical applications of meteorology 4
§ 4. General survey of meteorological research methods 5
§ 5. Organization of meteorological research in the USSR; the Hydrometeorological Service of the USSR 7
§ 6. Principal meteorological elements 9

Part One
GENERAL PROPERTIES AND STRUCTURE OF THE ATMOSPHERE

Chapter 2. THE COMPOSITION OF ATMOSPHERIC AIR 15
§ 1. Composition of dry air in the lower atmosphere 15
§ 2. Equation of state of dry air 18
§ 3. Equation of state of water vapor. The correlation between the various characteristics of atmospheric humidity 19
§ 4. Equation of state of moist air. Virtual temperature 21
§ 5. Water vapor in the atmosphere 23
§ 6. Ozone in the atmosphere 24
§ 7. Atmospheric aerosols 28
§ 8. Composition of air at high altitudes 33

Chapter 3. TURBULENT MIXING IN THE ATMOSPHERE 37
§ 1. Turbulent character of motions in the atmosphere 37
§ 2. Coefficient of turbulence 39
§ 3. Equation of turbulent mixing 44

Chapter 4. STRUCTURE OF THE ATMOSPHERE 46
§ 1. Methods of investigation of the upper atmospheric layers 46
§ 2. Stratification of the atmosphere 48
§ 3. Vertical extent of the atmosphere. The escape (dissipation) of gases from the atmosphere 51
§ 4. Hypotheses concerning the origin and evolution of the atmosphere 53
§ 5. Distribution of temperature, pressure and density in the upper atmospheric layers . . . 54
§ 6. Air currents in the atmosphere. General circulation of the atmosphere. Jet streams . . 59
§ 7. Horizontal inhomogeneity of the troposphere. Air masses and fronts 65

Chapter 5. DISTRIBUTION OF PRESSURE IN THE ATMOSPHERE 69
§ 1. Hydrostatic equation of the atmosphere 69
§ 2. Barometric formulas 70
§ 3. Complete barometric formula (Laplace formula) 74
§ 4. Simplified barometric formula (The barometric step) 76
§ 5. Applications of the barometric formula 78
§ 6. Applicability of the barometric formula to the upper atmospheric layers 79
§ 7. Mass of the atmosphere and its vertical distribution 80
§ 8. Factors influencing the variation of pressure 81

§ 9. Isobaric surfaces and isobars. Pressure regions 82
§ 10. Mean distribution of pressure at the earth s surface 84
§ 11. Geopotential. The barometric formula for the geopotential 87
§ 12. Absolute and relative geopotential. Charts of pressure topography 90
§ 13. Pressure distribution in the atmosphere at different levels 92
§ 14. Variation of pressure in time 97

Chapter 6. FUNDAMENTALS OF ATMOSPHERIC THERMODYNAMICS 99
§ 1. Fundamental equations 99
§ 2. Polytropic changes of state. The equation of polyttopy 101
§ 3. Adiabatic processes. The Poisson equation 102
§ 4. Dry adiabatic temperature gradient 103
§ 5. Potential temperature and its variation with height 104
§ 6. Adiabatic changes of state of moist air 107
§ 7. Moist adiabatic processes. The moist adiabatic gradient 108
§ 8. *Entropy and potential temperature* 111
§ 9. Dependence of the buoyancy of saturated water vapor on temperature (Clausius-Clapeyron equation) 113
§ 10. Condensation level 114
§ 11. Equivalent and equivalent-potential temperature. Pseudo-adiabatic processes 116
§ 12. Conditions of vertical stability of the atmosphere 118
§ 13. Thermodynamic graphs (diagrams) 124
§ 14. Energy of instability and its determination 125
§ 15. Determination of the instability by the layer method 127

Part Two
RADIANT ENERGY

Chapter 7. RADIANT ENERGY IN THE ATMOSPHERE 130
§ 1. The fluxes of radiant energy in the atmosphere 130
§ 2. Fundamental quantitative characteristics of the radiation field 131
§ 3. Fundamental laws of emission 135
§ 4. The sun and its radiation. The solar constant 137

Chapter 8. INFLUENCE OF THE ATMOSPHERE ON FLUXES OF RADIANT ENERGY 143
§ 1. Fundamental relations. The path of a solar ray in the atmosphere 143
§ 2. Integral attenuation of the flux of direct solar radiation in the atmosphere 146
§ 3. *Absorption of radiant energy in the atmosphere* 148
§ 4. Scattering of radiant energy in the atmosphere 152
§ 5. Atmospheric transmission and its characterization. The transmission coefficient. The turbidity factor 160
§ 6. Reflection of solar radiation. The albedo 163

Chapter 9. SHORT-WAVE RADIATION IN THE ATMOSPHERE 168
§ 1. Direct solar radiation at the earth's surface 168
§ 2. Direct solar radiation according to observations and calculations 172
§ 3. Diffuse radiation 176
§ 4. Total radiation 184

Chapter 10. LONG-WAVE RADIATION OF THE EARTH'S SURFACE AND ATMOSPHERE. THE RADIATION BALANCE 188
§ 1. Radiative properties of different regions of the earth's surface and atmosphere 188
§ 2. Approximate equations of transfer of radiant energy 191
§ 3. Emission of earth's surface and counter-emission of the atmosphere. Effective emission . 194
§ 4. Methods of calculating the effective emission 196
§ 5. Radiation balance 198

Part Three

THERMAL REGIME OF THE ATMOSPHERE AND EARTH'S SURFACE

Chapter 11. HEAT TRANSFER IN THE UPPERMOST LAYERS OF SOIL AND WATER. THERMAL REGIME OF SOIL AND WATER 204

§ 1. Propagation of heat in the soil. 204

§ 2. Thermal properties of the soil 207

§ 3. Heat transfer in the soil 209

§ 4. Daily and annual march of soil temperature 211

§ 5. Influence of plant cover and other factors on the thermal regime of the soil 214

§ 6. Snow cover and its significance for the thermal regime of the soil 215

§ 7. Freezing of soil. Permafrost 216

§ 8. Propagation of heat in water 218

Chapter 12. HEAT TRANSFER IN THE ATMOSPHERE 220

§ 1. Fluxes of heat in the atmosphere 220

§ 2. Influx of heat and its relationship with the flux 225

§ 3. Radiative and turbulent influx of heat 225

§ 4. Internal sources of heat 226

§ 5. Equation of heat influx 227

§ 6. Advection of heat and cold 228

Chapter 13. TEMPERATURE OF THE LOWER ATMOSPHERIC LAYERS 230

§ 1. Mean distribution of temperature at the earth's surface 230

§ 2. Annual march of air temperature 234

§ 3. Daily march of air temperature 235

§ 4. Vertical variation of temperature in the surface layer of air 238

§ 5. Frost 241

§ 6. Thermal convection in the atmosphere 243

Chapter 14. TEMPERATURE OF THE FREE ATMOSPHERE 245

§ 1. Mean temperature distribution at various levels in the troposphere 245

§ 2. Vertical variation of air temperature 247

§ 3. Annual and daily march of air temperature in the troposphere 249

§ 4. Height and temperature of the tropopause 250

§ 5. Inversions 252

§ 6. Theoretical considerations on the vertical distribution of temperature in the free atmosphere 257

Chapter 15. THERMAL BALANCE OF THE EARTH'S SURFACE AND ATMOSPHERE 262

§ 1. Equation of thermal balance for the earth's surface 262

§ 2. Equation of thermal balance for the atmosphere and earth-atmosphere system 264

§ 3. Relative magnitudes of the components of thermal balance. Daily and annual march . . 266

§ 4. Thermal balance of the earth as a whole 272

Part Four

MOISTURE REGIME OF THE ATMOSPHERE

Chapter 16. PHASE TRANSITIONS OF WATER 274

§ 1. Conditions of phase equilibrium 274

§ 2. Temperature dependence of buoyancy of saturated vapor over water and over ice . . . 276

§ 3. Thermodynamic diagram of phase equilibrium 280

Chapter 17. EVAPORATION 283

§ 1. Process of evaporation from the kinetic theory standpoint 283

§ 2. Special features of evaporation in nature 288

§ 3. Evaporation and evaporation capacity 289

§ 4. Equation of turbulent diffusion of water vapor 290

§ 5. Empirical formulas for the calculation of evaporation 294

§ 6. Results of observations of evaporation and evaporation capacity 295

§ 7. Distribution of water vapor in the atmosphere. Daily and annual march of atmospheric humidity 297

Chapter 18. FORMATION OF THE LIQUID AND SOLID PHASE OF WATER IN THE ATMOSPHERE 300
§ 1. Condensation and sublimation of water vapor 300
§ 2. Vapor pressure over the surface of droplets 302
§ 3. Formation and growth of embryonic droplets 306
§ 4. Condensation nuclei 308
§ 5. Supercooled droplets. Formation of ice crystals 311

Chapter 19. FOGS AND CLOUDS 314
§ 1. General information on fogs and clouds 314
§ 2. Aggregate state of cloud particles 316
§ 3. Microstructure of droplet fogs and clouds 317
§ 4. Size distribution of droplets in fog and cloud 319
§ 5. Microstructure of ice-crystal and mixed fog and cloud 327
§ 6. Processes leading to the formation of fog. Classification of fogs 331
§ 7. Physical foundations of fog forecasting 333
§ 8. Basic processes responsible for the formation of clouds 336
§ 9. Altitude and vertical extension of clouds 341

Chapter 20. PRECIPITATION 344
§ 1. Surface deposition and conditions of formation 344
§ 2. Characterization and classification of precipitation from clouds 346
§ 3. Size and fall velocity of particles of precipitation 347
§ 4. Evaporation of raindrops 349
§ 5. Growth of droplets in clouds 353
§ 6. Precipitation from droplet clouds 360
§ 7. Precipitation from ice and mixed clouds 361
§ 8. Amounts and global distribution of precipitation 363
§ 9. Hydrologic cycle. The moisture balance at the earth's surface 364
§ 10. Artificial stimulation of cloud and fog 367

Part Five
AIR MOTIONS

Chapter 21. GENERAL REGULARITIES OF ATMOSPHERIC MOTION 373
§ 1. Forces in the atmosphere 373
§ 2. Equations of motion 378
§ 3. Trajectories and streamlines 380
§ 4. Frictionless stationary motion 381

Chapter 22. WIND VARIATION WITH ALTITUDE 387
§ 1. Influence of friction on air motion 387
§ 2. Wind profile in the surface layer 389
§ 3. Wind profile in the boundary layer of the atmosphere 395
§ 4. Vertical variation of wind speed and direction in the free atmosphere. Thermal wind . 398

Chapter 23. AIR MOTION IN THE BOUNDARY LAYER OF THE ATMOSPHERE 402
§ 1. Daily march of wind speed and direction 402
§ 2. Structure of the wind in the lower layers of the atmosphere 404
§ 3. Influence of obstacles on wind 406
§ 4. Orographic winds. Foehn. Bora 409
§ 5. Local circulations. Breezes, mountain and valley winds 412

Part Six
OPTICAL PHENOMENA IN THE ATMOSPHERE

Chapter 24. OPTICAL PHENOMENA IN THE ATMOSPHERE DUE TO SCATTERING, ABSORPTION AND REFLECTION OF LIGHT 416
§ 1. Fundamentals of photometry with application to atmospheric optics 416
§ 2. Apparent form of the sky and related phenomena 419

§ 3. Luminance of the sky 420
§ 4. Blue color of the sky 422
§ 5. Polarization of light scattered by the atmosphere 424
§ 6. Daytime illuminance 429
§ 7. Twilight, dawn and dusk. Twilight illumination 431
§ 8. Nighttime illumination. Illumination due to moonlight. Nightglow 438
§ 9. Luminance, color and reflectance of objects of the natural landscape 441

Chapter 25. VISIBILITY 444
§ 1. Visual range. Principal factors governing visual range 444
§ 2. Effect of the atmosphere on the apparent luminance of an object. The air light equation . 445
§ 3. Luminance contrasts. Visual acuity 446
§ 4. Fundamental theorems of the theory of visual range 449
§ 5. Visual range of real objects 452
§ 6. Visibility of light sources 454
§ 7. Visibility in fog and rainfall zones 455

Chapter 26. REFRACTION OF LIGHT IN THE ATMOSPHERE 457
§ 1. Equation of the trajectory of a light ray in the atmosphere 457
§ 2. Astronomical and terrestrial refraction 461
§ 3. Phenomena due to astronomical refraction 463
§ 4. Lifting and lowering of horizon. Anomalous refraction 464
§ 5. Mirage 467

Chapter 27. OPTICAL PHENOMENA DUE TO THE PRESENCE OF WATER DROPLETS AND ICE CRYSTALS IN THE ATMOSPHERE 470
§ 1. Rainbows 470
§ 2. The halo 474
§ 3. Coronas, glories and other analogous phenomena 478

Part Seven
ELECTRICAL PHENOMENA IN THE ATMOSPHERE (ATMOSPHERIC ELECTRICITY)

Chapter 28. STATE OF IONIZATION OF THE ATMOSPHERE 483
§ 1. Fundamental concepts. Ions in the atmosphere and electrical conductivity 483
§ 2. Principal ionizing agencies in the atmosphere 487
§ 3. Ion destruction processes. Ion recombination 490
§ 4. Ion concentration and conductivity of the atmosphere according to observational data . . 493

Chapter 29. THE IONOSPHERE 500
§ 1. Ionosphere. Formation of the ionized layer 500
§ 2. Propagation of electromagnetic waves in the ionosphere 503
§ 3. Structure of the ionosphere according to observational data 505

Chapter 30. THE ELECTRIC FIELD IN THE ATMOSPHERE 508
§ 1. Fundamental relations 508
§ 2. Electric field strength according to observational data 509
§ 3. Daily and annual march of the electric field strength 512
§ 4. Correlations between electrical field strength and other electrical and meteorological elements of the atmosphere 515

Chapter 31. THUNDERSTORM ELECTRICITY 518
§ 1. Electric charges of precipitation 518
§ 2. Distribution of charges in thunder clouds 521
§ 3. Thunderstorm discharges. Lightning and the lightning mechanism 524
§ 4. Global distribution of thunderstorm activity 528
§ 5. Silent discharges 529
§ 6. Atmospherics 531
§ 7. Electric current balance in the atmosphere 532

Chapter 32. POLAR AURORAS 536
§ 1. Geographic distribution of auroras and their forms 536
§ 2. Height of polar auroras. Their distribution in space. Periodic variations 539
§ 3. Color, intensity and spectrum of polar auroras 540
§ 4. Theory of polar auroras 541

Part Eight
ATMOSPHERIC ACOUSTICS

Chapter 33. SPEED OF SOUND IN THE ATMOSPHERE 545
§ 1. Speed of sound in an immobile homogeneous medium 545
§ 2. Dependence of the speed of sound in the atmosphere on temperature and humidity . . . 546
§ 3. Speed of sound in the presence of wind 548

Chapter 34. PROPAGATION OF SOUND IN THE ATMOSPHERE 550
§ 1. Path of a sound ray in the atmosphere 550
§ 2. Reflection, refraction and absorption of sound in the atmosphere 554
§ 3. Utilization of observations on sound propagation for research into the upper atmosphere . 556

APPENDICES 559

BIBLIOGRAPHY 561

Annotation

The book presents an exposition of a course in meteorology (physics of the atmosphere) in conformity with the syllabus approved for hydrometeorological institutes and universities.

On the basis of the latest researches, the book systematically presents information on the structure and properties of the atmosphere, and the physical processes and phenomena taking place in it.

The book is intended as a textbook for hydrometeorological institutes and universities. It can serve as a manual for correspondence courses and also for practical workers in the field of meteorology.

PREFACE

Over ten years ago a manual entitled "A Course in Meteorology (Physics of the Atmosphere)" was published by a group of lecturers including A. S. Zverev, B. V. Kiryukhin, K. Ya. Kondrat'ev, E. S. Selezneva and M. I. Yudin under the direction and with the participation of the present author. However, the problem of producing a textbook on this course remained unsolved.

The present book is intended as a textbook for students of the meteorological specialities at hydrometeorological institutes, universities and other educational institutions. Fundamental data relating to meteorology are presented systematically in conformity with the syllabus approved for hydrometeorological institutes.

Our knowledge has increased considerably in the last ten years in a number of branches of meteorology, especially in connection with the extensive atmospheric researches carried out during the International Geophysical Year (1957–1958). New data have been obtained on the composition and structure of the upper atmosphere, and on meteorological processes in the equatorial zone, over oceans, in the Arctic and the Antarctic. Resourceful studies have been carried out in cloud physics and important advances made towards the active stimulation of cloud and fog. Theoretical investigation of atmospheric processes has developed apace.

The author has sought to incorporate the salient results of these researches in the book.

The book is based on the above-mentioned "A Course in Meteorology" and on lectures delivered by the author over a period of many years at the Faculty of Physics of the Leningrad State University and by N. P. Tverskaya at the Leningrad Hydrometeorological Institute.

The part entitled "Moisture Regime of the Atmosphere" was written in collaboration with B. V. Kiryukhin.

The author expresses his gratitude to his colleagues at the Main Geophysical Observatory and departments of atmospheric physics of the Leningrad and Moscow State Universities, who participated in discussion of the manuscript and contributed critical remarks, and also to V. A. Gavrilov, V. P. Kolokolov, K. Ya. Kondrat'ev, M. P. Timofeev and M. I. Yudin, who read certain parts prior to publication.

The author is greatly indebted to E. S. Selezneva, who took upon herself the task of editing the book and gave invaluable assistance in the final preparation of the book for printing in view of the author's illness.

Chapter 1

INTRODUCTION

§ 1. The subject-matter of meteorology

The earth is surrounded by a gaseous envelope called the atmosphere.

The earth's atmosphere, its structure, properties and the physical processes and phenomena taking place within it are studied from all points of view in meteorology.

The term "meteorology" itself is an ancient one. As far back as the fourth century B.C., the Greek philosopher Aristotle produced a work entitled "Μετεωρολογια," which, translated literally, means the science (Greek λογοσ) of atmospheric phenomena (μετεωρα). All information relating to phenomena observed in the atmosphere was collected in this work; an attempt was even made, in fact, to explain them in a philosophical, speculative manner. However, there was still a long way to go from there to the birth of meteorology as a genuine natural science. This occurred much later, after the Dark Ages (17th century), when the medieval scholastic outlook had been replaced by the new scientific method of research into natural phenomena, namely, that based on exact experiment.

The practical needs of society (i.e., primarily those of industry and trade), coupled with the invention of the earliest meteorological instruments (thermometer at the end of the 16th century and barometer in 1643) which made instrumental observation possible, gave rise to the development of meteorology. On the basis of the mechanical and physical laws then being discovered, the first attempts were made to explain certain meteorological phenomena and generalize the vast uncoordinated mass of data which had accumulated. Initially the science of meteorology developed as a part of physics, and it was only in the first half of the 18th century that, without losing its links with physics, it emerged as an independent discipline.

Nearly all phenomena and processes occurring in the atmosphere form and develop under the direct or indirect influence of the energy delivered to the earth by the sun.

Owing to the complex processes of conversion of solar energy into other forms of energy, the atmosphere is the site of air-mass movements and of a wide variety of thermodynamic processes; a constant exchange of heat and moisture takes place between individual regions of the atmosphere. As a result, changes in atmospheric pressure, wind, storms, air temperature variations, cloud formation, precipitation and so forth are observed. Further, various optical, electrical, acoustic and other phenomena also take place in the atmosphere.

It should be stressed that, rather than developing in isolation, atmospheric phenomena are very closely linked not only to other atmospheric phenomena but also to processes occurring at the earth's surface – in the surface layer of the crust and in the hydrosphere. Owing to the close connection between the latter and atmospheric phenomena, some of these surface processes are among the questions treated in meteorology.

By studying phenomena in the atmosphere meteorology seeks to establish their interrelations as well as their relationships with other natural phenomena. The principal aim of meteorology, like that of any other science, is to study the laws of nature. It cannot, however, confine itself to observing, describing and explaining atmospheric phenomena. A progressive science must also satisfy the practical requirements of man; its deductions should be applied for the benefit of society and its ultimate goal should be to find practical ways of controlling natural phenomena.

Accordingly, the tasks facing meteorology can be summarized as follows:

1) to obtain accurate factual data characterizing the atmosphere and the phenomena observed in it, and to describe these qualitatively as well as quantitatively;

2) to find, from the analysis of observed data, the correct explanation for atmospheric phenomena and to establish the laws governing their development;

3) to evolve methods for accurate forecasting of the development of atmospheric processes on the basis of these laws;

4) to exploit the laws governing the development of atmospheric processes for harnessing the forces of nature.

§2. Relationship between meteorology and other sciences; subdivisions of meteorology

Meteorology is intimately linked to other scientific disciplines, being concerned with many problems examined from different standpoints in other sciences. Moreover, meteorology makes extensive use of results obtained by related sciences, and its deductions and data are in turn widely employed beyond its confines.

Meteorology is directly related first of all to geophysics, which comprises the complex of sciences dealing with the physical properties of the earth and the physical processes taking place on it. In geophysics the earth is regarded as having three envelopes: the lithosphere, hydrosphere and atmosphere. Meteorology studies the physical phenomena occurring in one of these envelopes and may thus be regarded as a fundamental branch of geophysics.

So far as the nature of its problems and the methods used are concerned, meteorology or the physics of the atmosphere is a physical science and can develop successfully only if it makes the maximum use of all the achievements of physics. Its closest links are with such branches of physics as mechanics and aerodynamics, optics, electricity, heat and so on.

On the other hand, like all geophysical disciplines, meteorology is closely connected to certain geographical sciences. This is because atmospheric processes, linked as they are to processes at the earth's surface, take somewhat different forms under different physico-geographic conditions.

The connection between meteorology and geography is particularly apparent in the fact that meteorology devotes a great deal of attention to explaining the general patterns of the geographic distribution of meteorological phenomena, as well as to studying the long-range weather regime, i.e., climate, characteristic of each geographical region. The science of climate - climatology - is one of the most important subdivisions of meteorology.

The geographic situation must also be considered in solving the paramount practical problem of meteorology, that of forecasting the weather. This problem is one which has been facing meteorology since the earliest days of its development; it is also the problem which has always received most attention. How important a role weather plays in the life and activities of mankind, and how important it is to foresee its changes, is well known to all.

By making extensive use of results on the physical aspects of atmospheric processes and applying special methods of complex analysis to phenomena occurring over large territories (which may be the entire earth), meteorology is finding the solutions to the problem of weather forecasting. The study of the laws of development of atmospheric processes over large geographic areas for the purpose of predicting weather and evolving the methods required for such prediction constitutes the subject matter of a large and fairly independent branch of meteorology known as synoptic meteorology.

Meteorology is related not only to physics, geophysics and geography but also to other sciences such as astrophysics. The radiation from the sun, the connection between processes taking place in the sun and in the earth's atmosphere, the question of the origin and structure of the atmosphere and many other questions are common to atmospheric physics and astrophysics. Similar ties can be shown to exist between meteorology and general astronomy, geodesy, physical chemistry and other sciences.

Meteorology is moreover closely linked to a whole series of applied disciplines to which it supplies necessary data.

The extensive spectrum of processes studied in meteorology may be divided into three groups according to the degree of influence exercised on their development by the earth's surface. This influence is unquestionably strongest in the relatively thin layer immediately above the earth's surface; this layer of the atmosphere, the so-called surface layer, * is estimated to be a few tens of meters thick. Researches into this layer and processes occurring within it have evolved in recent years into the special branch of meteorology called physics of the surface layer of the atmosphere. The development of studies in this field are of exceptionally great practical significance, since it is basically in this layer that all plant and animal life (including human) develops.

At the beginning of this century the study of processes taking place at higher levels in the atmosphere (the so-called free atmosphere), where the direct influence of the underlying ground is distinctly weaker, was singled out in a special branch of meteorology called aerology, or, as it is now frequently called, free-atmosphere physics.

The study of the uppermost layers of the atmosphere is the subject of a branch of meteorology which has developed very intensively in recent years; this branch is called upper-atmosphere physics.

* [Often called the "lower atmosphere" in Western usage. The latter term is used in Russian in a different meaning. See p. 50.]

A characteristic feature of these layers is the high degree of rarefaction of the gases; the action of cosmic factors on the atmosphere is especially strong here, while the influence of the earth's surface vanishes. Of late, in addition to their general scientific value upper-atmosphere investigations have taken on tremendous practical significance in connection with rocketry, the launching of artificial satellites and the question of interplanetary travel.

As meteorology developed there gradually emerged a series of branches differing in the phenomena and processes they studied and also in the methods they employed. The number of such branches has been increasing steadily and some of these have now assumed the character of more or less independent specialized disciplines. To mention only a few, the study of the mechanism and peculiarities of atmospheric processes, chiefly by theoretical methods, making extensive use of theoretical physics, fluid mechanics and thermodynamics constitutes dynamic meteorology. The comprehensive study of the radiant energy of the sun and its transformations in the atmosphere is the concern of actinometry. Closely related to the latter is atmospheric optics, which is the study of the numerous optical phenomena in the atmosphere. The electrical properties of the atmosphere and electrical phenomena observed in it are studied in a branch called atmospheric electricity.

Owing to the important role of clouds and the various forms of precipitation emanating from them, the study of the complex of processes leading to the formation of clouds and precipitation, and also the development of methods influencing the latter, has been singled out in a branch called the physics of clouds and precipitation.

Finally, in meteorology, a great deal of attention is devoted to the development of instruments and to questions relating to techniques of meteorological observation. These questions are discussed in specialized textbooks and manuals, and we will not be concerned with them here.

§3. Practical applications of meteorology

The atmosphere is the environment in which we spend our entire life. It is therefore rather natural that results obtained by meteorology have many and far-reaching practical applications.

To a greater or lesser extent, weather and climate and a host of special atmospheric phenomena affect all aspects of human activity. Where the laws governing atmospheric processes are known they can be utilized not only to take preventive measures against unfavorable changes of weather but also to fight actively against these changes.

The importance of meteorological factors in agriculture is universally known. The growth, development and, ultimately, yield of agricultural crops depend on the amount of heat, light and moisture they receive at different stages in their development. With the help of weather forecasts it is possible to plan ahead for various agricultural operations and to take early measures against such destructive phenomena as frost and drought.

Meteorological data are no less important for aviation. Successful flights require accurate evaluation of the actual state of the weather and expected weather changes. For, visibility, clouds (especially thunder clouds), high air-stream velocities and many other phenomena in the

atmosphere have a very great influence on aviation. This calls for thorough study of the ocean of air all the way up to the highest levels.

The significance of weather conditions for sea and river transportation is common knowledge; it is equally great for rail and other overland transportation. The successful functioning of all forms of land transportation is often hampered by such phenomena as gales, snowstorms, showers, snow drifts, glaze frost, fog and so forth.

Telegraph and telephone lines, as well as high-voltage wires, are also vulnerable to the influence of atmospheric phenomena: formation of heavy glaze on wires often causes breakage, the functioning of communication lines and electric wires is often interrupted during thunderstorms, and so on.

Radio communication is also dependent on the state of the atmosphere; in particular, the propagation of radio waves over long distances takes place exclusively as a result of the ionization of the upper layers of the atmosphere. Radio communication is frequently disrupted or disturbed by atmospheric obstacles due to such phenomena as thunderstorms.

Aside from those already mentioned, nearly all other branches of the economy (industry, construction, health services, etc.) make use of meteorological data to some extent.

Owing to the vast range of practical applications of meteorological data, individual problems of greatest value for a given practical field were selected at an early date and subjected to thorough investigation in relation to the specific requirements of the field. Although "there are no pure and applied sciences – there are sciences and their applications" (Academician A.N. Bakh), one frequently speaks of applied meteorology. In this respect agricultural meteorology, and also aviation meteorology, are the most extensively developed. Attempts have been made to single out many other "secondary meteorologies," e.g., maritime, transport, forest, military and medical meteorology.

Russian meteorological scientists, beginning with the founder of Russian science M.V. Lomonosov, were always guided in their work by the interest of the state and people. The link between meteorology and practical needs became particularly evident in this country subsequent to the October Revolution; this in turn explains the rapid growth of Soviet meteorology.

§4. General survey of meteorological research methods

The methods of investigation used in meteorology, as in other geophysical disciplines, are chiefly physical.

Meteorology draws the major part of the factual material for its deductions from observations under natural conditions. Side by side with very simple observations, performed with very simple instruments (thermometer, barometer, psychrometer) or none at all, many special observations involve highly complicated and refined physical equipment as well as aircraft, rockets, electronic equipment and so on. An essential feature of meteorological processes and phenomena is that they cover extremely large areas and change drastically in time; the place and time of their inception, moreover, are usually not known in advance. This necessitates organizing the observations (as in all geophysical investigations) in such a way as to permit early detection of the onset of

any phenomenon and constant tracking of its development over the entire region it covers.

Uninterrupted observation in space and time can be achieved by organizing a network of observation points distributed at sufficiently close intervals over the entire surface of the globe with observations regularly performed with standard instruments according to a unified program.

There are now many thousands of such observation points (stations) scattered over the globe. The network of meteorological, and also aerological, stations has expanded particularly rapidly; their function comprises observation of the more important atmospheric phenomena and regular communication of the principal quantitative and qualitative characteristics of the state of the atmosphere. Such characteristics are known as meteorological elements (see § 7). Special observatories and research institutes deal with the detailed physical investigation of individual phenomena, including research into particularly complex phenomena with the aid of special installations and, frequently, highly complex equipment. Examples are researches into thunderstorm processes, optical and acoustic phenomena in the atmosphere, microphysical characteristics of clouds and precipitation (including methods of controlling them), and so forth. Observation of upper atmosphere phenomena is particularly complex and requires advanced technology.

In addition to observations carried out at permanent stations and observatories, extensive use is made in meteorology of the "expedition method" of investigation, which permits study of inaccessible regions and investigation of particular phenomena under optimum conditions of observation. Examples are the expeditions to mountain districts and into the interior of the Arctic and Antarctic.

Quite apart from observations in natural conditions, the significance of experimental research in meteorology is increasing every year.

In its original sense, which presupposes artificial reproduction of the phenomenon under investigation, experiment is very difficult to achieve under natural conditions. It is impossible to reproduce artificially in a natural situation such phenomena as aurora polaris, thunderstorms and squalls. We are thus confined for the main to laboratory investigations of the phenomena of interest and their individual aspects, chiefly through the agency of models. The greatest difficulties encountered in setting up experiments stem from the necessity of ensuring the similitude of the laboratory models to their originals in nature.

True, experimental investigations are also carried out under natural conditions in connection with the artificial control of certain meteorological phenomena; such, for instance, are the successful experiments with frost and drought control, artificial influence on fog, clouds, precipitation and certain other phenomena. However, for the most part these experiments are still being conducted on a small scale.

A variety of special methods is employed in meteorology in order to generalize the extensive material yielded by observations and draw conclusions of a scientific and practical nature. We mention merely that wide use is often made – especially in climatology – of the methods of mathematical statistics. Special graphical methods of processing and analysis of materials are also of considerable significance.

In view of the extreme complexity of most meteorological phenomena, statistical methods are helpful in establishing empirical regularities. Detailed studies of phenomena and accurate quantitative calculations would be impossible in the absence of a mathematical theory of these phenomena based on the general laws of physics. Attempts to create such theories for a number of phenomena were made at an early date. However, it is only at the present time that the extensive store of factual material relating to processes in the atmosphere has made it possible to attempt their study on the basis of the general laws of physics, making extensive use of the methods of mathematical analysis.

These quantitative theories, carefully verified by observations, have led to a deeper and more comprehensive understanding of the physical basis of meteorological phenomena and have opened up approaches for further experimental investigation. However, many difficulties arise in the construction of an exact theory, due primarily to the complexity of the phenomena in question and to the necessity of allowing for many factors not always sufficiently well known. As a result the equations with which one has to deal in theoretical investigations are very involved, making it difficult ot obtain general solutions and frequently necessitating considerable simplifications. Achievements in this field are nonetheless very considerable; the theories that have been evolved provide a sufficiently accurate expression of the relationships observed in nature and can be used for practical calculations.

It is obvious from the above-said that the various methods of research current in meteorology are all basically physical. The principal ones are:

1) method of observation under natural conditions;
2) method of experimentation;
3) method of theoretical analysis of the phenomena on the basis of the laws of physics and mechanics with the aid of a suitably adapted mathematical apparatus.

§ 5. Organization of meteorological research in the USSR; the Hydrometeorological Service of the USSR

The importance and economic value of the questions treated in meteorology, and also the extent of their practical applications in all spheres of human society, have been responsible for the great expansion of meteorological research. All countries in the world have state-managed institutions dealing with the organization of meteorological observations and their practical ulilization. In the USSR, where science is directed toward the solution of problems of general economic significance, meteorological research is a concern of national importance. Lenin pointed this out in June 1921 when he signed the decree "On the organization of a meteorological service in the RSFSR". Later resolutions of the government of the USSR dealt with problems relating to the most efficient organization of work in meteorology and provided great opportunities for successful development.

The needs of the national economy in the sphere of meteorology are served in our country by GUGMS,* under the Council of Ministers of the USSR, and by the subordinate Regional and Republic Boards of the Hydrometeorological Service.** In all its work the GUGMS relies on a series of geophysical observatories and research institutes incorporated in its system. Of these we might mention first of all the oldest scientific institution in the field of meteorology, the Voeikov GGO, † and also such great scientific centers as the TsIP †† and TsAO ‡.

A large volume of scientific work on crucial meteorological problems is being carried out at the Institute of Physics of the Atmosphere (Institut fiziki atmosfery) and Institute of Applied Geophysics (Institut prikladnoi geofiziki) of the Academy of Sciences of the USSR), and also at the Academies of science of the union republics. Meteorological research is likewise being conducted in the universities and special hydrometeorological institutes, as well as in numerous other departments dealing with narrower problems connected with meteorology. Research institutes connected with the latter have sections working in the field of meteorology, examples being the Arctic and Antarctic Research Institute and the All-Union Plant-Growing Institute.

To ensure uniformity of the observations and the possibility of comparing data obtained at different stations in the network, the GUGMS publishes instructions and directives covering the execution of observations, the installation of instruments and their care, and also the question of the procedures to be followed in the recording and processing of observations. The instructions sanctioned by the GUGMS must also be followed by meteorological stations in other departments. These questions are all discussed in detail in special textbooks devoted to methods of processing and observation.

In many cases the investigation of meteorological phenomena creates a need for coordinated data covering several countries, as atmospheric processes develop over territories occupied by several states. This involves the exchange of results of current observations between countries and a unified program of observations, requirements fulfilled through the International Meteorological Organization under the aegis of the United Nations, of which the Soviet Union is a member.

The solution of many of the most important problems of meteorology requires especially complex, well-coordinated observations over large territories or even the entire globe. To deal with these requirements such measures as the First and Second International Polar Years (1882–1883, 1932–1933), and also the International Geophysical Year (1957–1958), have been instituted at various times by common agreement among a majority of countries; in the case of the IGY a great part of the work program was devoted to the solution of meteorological problems, and the USSR took a very active part in the operations.

* [Glavnoe upravlenie gidrometeorologicheskoi sluzhby — Central Board of the Hydrometeorological Service]

** [In abbreviated form — UGMS]

† [Glavnaya geofizicheskaya observatoriya im. A.I. Voeikova — Main Geophysical Observatory im A.I. Voeikov.]

†† [Tsentral'nyi institut prognozov — Central Weather Institute]

‡ [Tsentral'naya aerologicheskaya observatoriya — Central Aerological Observatory]

§6. Principal meteorological elements

The quantitative and qualitative characteristics of the state of the atmosphere which are obtained as a result of direct observations at the network of meteorological stations are called, as we saw earlier, meteorological elements. Their number is fairly large and is constantly increasing with the growth of demands imposed on meteorological observations. In later chapters many of these characteristics will be considered in detail. Here we confine ourselves to listing those which will crop up constantly starting from the earliest pages; we also indicate the units of measurement.

1. Air temperature is usually expressed in meteorology in degrees of the Celsius [centigrade] scale (t°C); the absolute scale of temperatures (T°K) is often used in theoretical calculations. It is well known that

$$T°\,\mathrm{K} = (273.16° + t)\,\mathrm{C} \approx 273(1 + \alpha t),$$

where α is the coefficient of volumetric expansion of gases, equal to

$$\frac{1}{273} = 0.003667.$$

2. Atmospheric pressure (air pressure) is the force of hydrostatic pressure of air acting on a unit area (1 cm^2). In immobile air it is equal to the buoyancy of air and is balanced at every point by the weight of a vertical column of air of unit cross section (1 cm^2) lying over the level under consideration.

The unit of pressure measurement in the CGS system is the bar, equal to 1 dyne/cm^2. However, as this unit is very small for meteorological computations, the term bar in meteorology is applied to a unit 10^6 times larger,* and the basic pressure unit is the millibar (1 mb = 10^{-3} bar = = 10^3 dyne/cm^2). In practice pressure is usually measured in terms of the height of a column of mercury in a barometer, expressed in millimeters (mm Hg). But the height of the mercury column in a barometer, as is well known, depends not only on pressure but also on the temperature of the mercury, as well as on the acceleration of gravity at the point of observation, which varies with elevation above sea level and latitude. The readings of a mercury barometer must therefore be reduced to a standard temperature (usually 0°C), and a standard altitude and latitude. The convention is to reduce readings to sea level and lat. 45°, and the normal (standard) pressure is taken to be that of a column of mercury 760 mm high with a base of 1 cm^2, at temperature 0°C and latitude $\varphi = 45°$, where the acceleration of gravity $g_0 = 980.6$ cm/sec^2.

Let us find the relation between millibars and millimeters of a mercury column. Since the density of mercury at 0°C is $\rho_0 = 13.596$ g/cm^3, for the force p_0 corresponding to normal pressure (760 mm Hg) we have

$$p_0 = 76\rho_0 g_0 = 76 \times 13.596 \times 980.6 = 1\,013\,250\ \mathrm{dyn/cm^2} = 1013.2\ \mathrm{mb}.$$

It is easy to compute from here that to a pressure of 1 bar = 1000 mb corresponds a mercury column of height 750.08 mm and that 1 mm Hg $\approx$ ≈ 1.333 mb $\approx \frac{4}{3}$ mb, while 1 mb ≈ 0.75 mm Hg $\approx \frac{3}{4}$ mm Hg.

* The meteorological bar is equal to 10^6 physical bars, a point which should always be borne in mind.

3. Atmospheric humidity. The following quantities are employed to characterize the humidity of air:

a) buoyancy (partial pressure) e of the water vapor contained in the air. Like air pressure the buoyancy of air is expressed in millibars or millimeters of a mercury column;

b) absolute humidity a, by which is meant the amount (mass) of water vapor contained in a unit volume: it is measured in grams per cubic meter (g/cm^3) and is equal to the density of water vapor (g/cm^3) multiplied by 10^6;

c) specific humidity q, which is the amount of vapor, expressed in grams, contained in one gram or one kilogram of moist air (g/g) (in practice usually g/kg). The specific humidity is therefore a pure number which gives the ratio of the mass of water vapor to the total mass of moist air;

d) mixture ratio s. This is the ratio of the mass of water vapor to the mass of dry air in a given volume of moist air. It is also expressed in g/g or g/kg;

e) relative humidity f, which is the ratio of the pressure e of the vapor contained in the air to the pressure E of saturated vapor at the same temperature; it is expressed in percents, i.e.,

$$f = \frac{e}{E} 100\%.$$

We recall that by the saturated vapor pressure E is meant the limiting pressure of the vapor at a given temperature in equilibrium over a flat surface of pure water;

f) moisture deficit d, which is the difference between E and e at the given temperature, i.e.,

$$d = E - e;$$

g) dew point τ, the temperature to which the air must be cooled (at constant pressure) in order for the vapor contained in it to reach the state of saturation.

The quantities which characterize atmospheric humidity are interrelated and can be expressed in terms of each other, as will be shown in § 4 of Chapter II. This is very important, since only some may be measured in practice.

4. Wind velocity and direction. Wind is the name applied to the movement of air with reference to the earth's surface. Since the horizontal component of this movement is considerably larger than the vertical one, it is usually the one recorded in observations. Wind direction is specified by the point on the horizon from which the wind blows. In meteorological observations the direction is named according to the points of the compass (8 or 16 points). Occasionally wind direction is specified by the azimuth in degrees east of north; thus a northerly direction corresponds to 0° (or 360°), an easterly direction to 90°, a southerly direction to 180° and a westerly direction to 270°.

Wind velocity is usually measured in m/sec or km/hr. In addition it is sometimes given in numbers on the Beaufort scale, derived from maritime practice (Table 1). Wind velocity expressed in numbers of the Beaufort scale is often called wind force.

In detailed studies the structure of wind, i.e., the variations of velocity and direction due to irregular motions of the air superposed on the mean air-mass motion, is also investigated. The phenomenon itself is called gustiness.

TABLE 1

Conventional units of wind velocity*

Description	Number	Velocity	
		m/sec	km/hr
Calm	0	0–0.5	0–1
Light air	1	0.6–1.7	2–6
Light breeze	2	1.8–3.3	7–12
Gentle breeze	3	3.4–5.2	13–18
Moderate breeze	4	5.3–7.4	19–26
Fresh breeze	5	7.5–9.8	27–35
Strong breeze	6	9.9–12.4	36–44
Moderate gale	7	12.5–15.2	45–54
Fresh gale	8	15.3–18.2	55–65
Strong gale	9	18.3–21.5	66–77
Whole gale	10	21.6–25.1	78–90
Storm	11	25.2–29.0	91–104
Hurricane	12	> 29	> 104

* [Beaufort scale.]

5. Cloudiness. Clouds form as a result of condensation of the water vapor contained in the atmosphere. The aggregate of clouds observed in the sky constitutes cloudiness. Cloud observations cover: the amount of clouds, i.e., cloudiness, their type, their altitude, and their direction and rate of motion. Cloudiness is estimated according to a ten-point scale and refers to the amount of sky covered by clouds; 0 corresponds to a cloudless sky, 10 to a completely overcast sky.

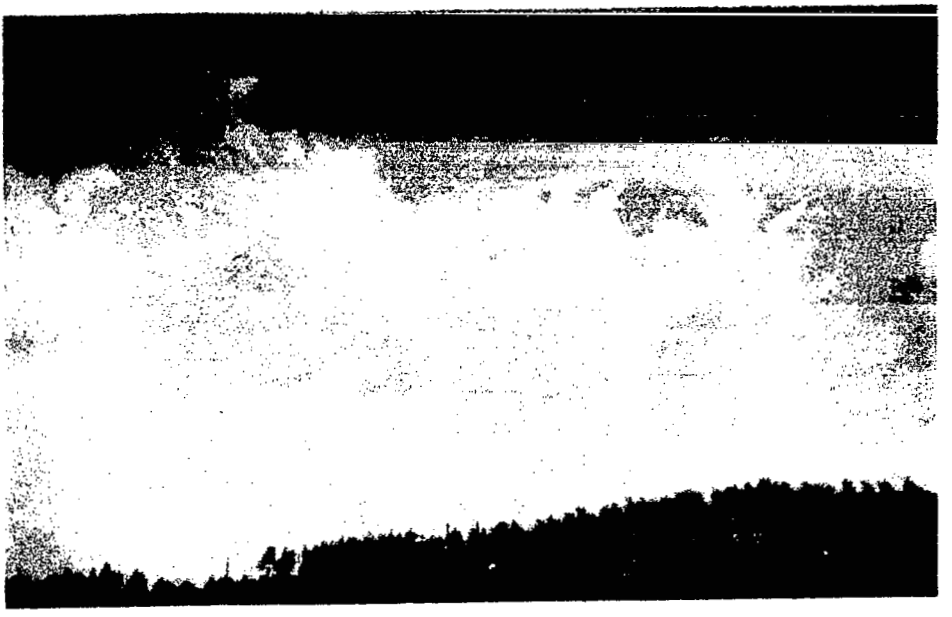
Cloud observations include determination of their type and appearance. We will pursue this question in detail later on; for the present we confine ourselves to giving the universally used morphological classification. It is based on two criteria: altitude and appearance (morphology).

In this classification clouds are labelled according to the altitude of their base as follows: high (above 6 km), middle (2–6 km) and low (below 2 km). In addition, a special group includes clouds of vertical development, which develop over a considerable altitude range; their base usually lies at a low level (below 2 km), while their apex is frequently high.

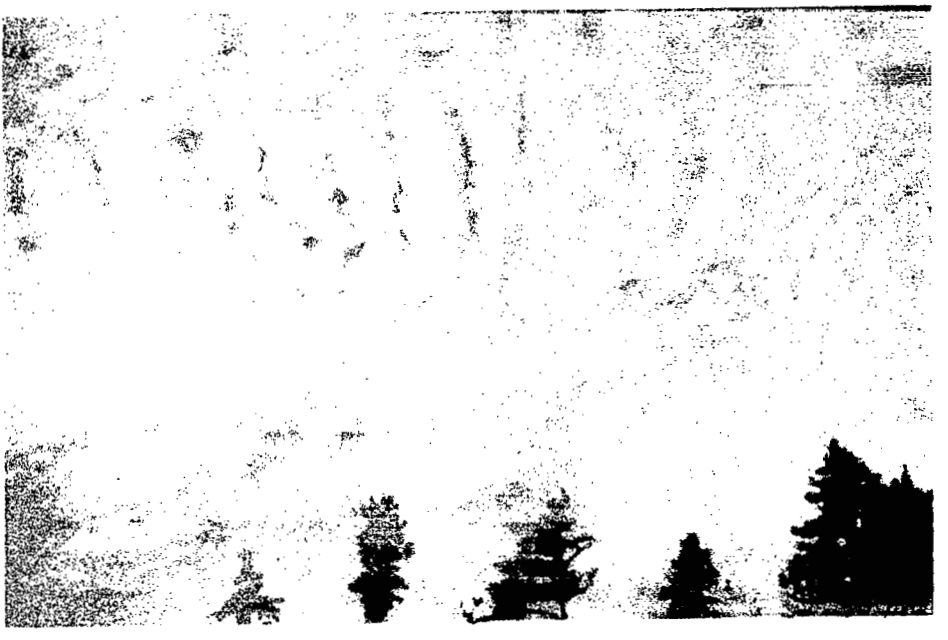
An especially complex aspect of cloud observations consists of assigning clouds to one of the numerous species distinguished in the morphological classification. This task can be made easier by employing the Cloud Atlas (Atlas oblakov), which provides descriptions and characteristic photographs. *

Each cloud type has its own name in Russian; in addition, Latin names are used internationally.

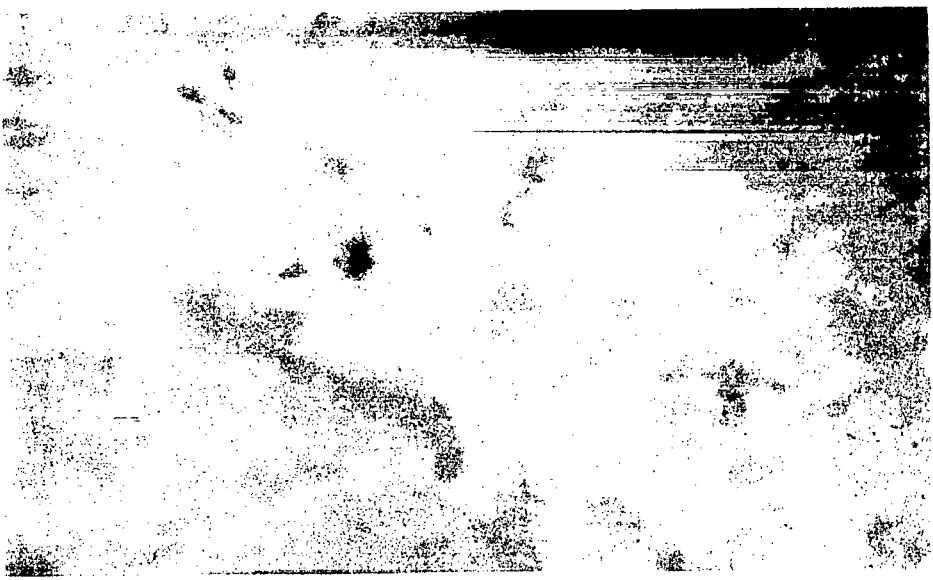
* A new Cloud Atlas was published in the USSR in 1957.



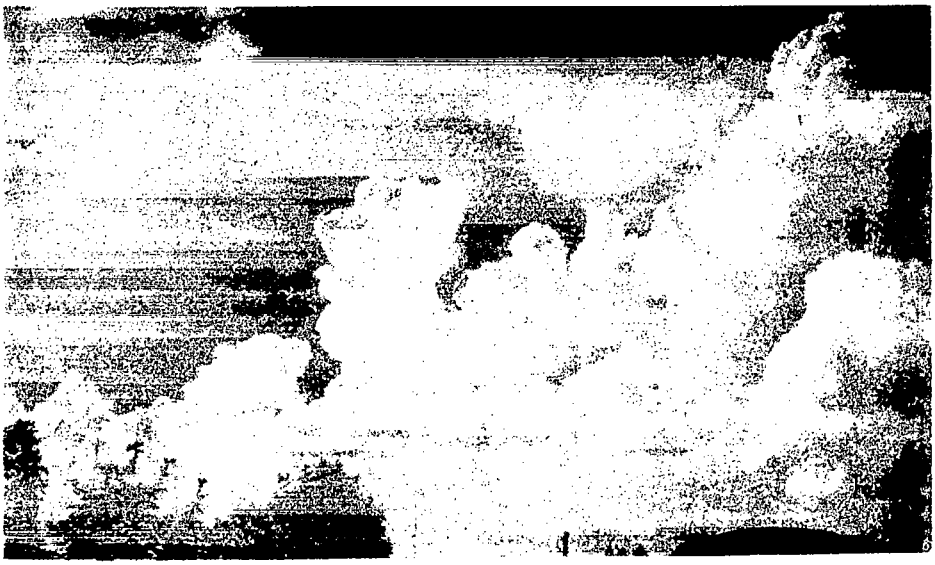
Cirrus clouds (Ci)



Altocumulus clouds (Ac)



Stratocumulus clouds (Sc)



Cumulus congestus clouds (Cu cong)

Without going into details we note that the grouping of clouds according to general appearance is as follows: (1) cumuliform types (separate detached cloud masses); (2) stratiform types (uniform sheet of clouds, forming a continuous ceiling when developed); (3) undulatus types (lower surface of clouds presents torn appearance, produces pattern composed of billows, filaments, flakes or bulges).

Clouds differ in density and the size of the external elements according to the altitude at which they are observed. Thus at low altitudes wave clouds appear as large billows, at high altitudes as fine ripples; a continuous sheet of high clouds is reminiscent of fine white veils, low clouds of the same type are similar to a thick gray curtain. Heavenly bodies often shine brightly through high clouds but are completely invisible through a low cloud layer.

The complete classification of clouds in use in the USSR is given in tabular form in Appendix II.

Table 2 lists the principal cloud types and their symbols.

TABLE 2

Cloud classification according to level and principal types

International name	Symbol	International name	Symbol
A. High clouds (base above 6 km)		C. Low clouds (base below 2 km)	
Cirrus	Ci	Stratocumulus	Sc
Cirrocumulus	Cc	Stratus	St
Cirrostratus	Cs	Nimbostratus	Ns
B. Middle clouds (base above 2–6 km)		D. Clouds of vertical development	
Altocumulus	Ac	Cumulus	Cu
Altostratus	As	Cumulonimbus	Cb

Part One

GENERAL PROPERTIES AND STRUCTURE OF THE ATMOSPHERE

Chapter 2

THE COMPOSITION OF ATMOSPHERIC AIR

Atmospheric air is a mixture of various gases, the principal ones being nitrogen (N_2), oxygen (O_2), argon (A), carbon dioxide (CO_2) and water vapor (H_2O). The other atmospheric gases occur in insignificant amounts (of some only "traces" can be found) and may be disregarded when studying the physical properties of air and the fundamental gas laws.

Suspended solid and liquid particles of natural and industrial origin (dust, smoke, products of vapor condensation, etc.) are always present in large amounts in the atmosphere; together they constitute the dispersed system called atmospheric aerosol.

If the air were purified of aerosol particles and water vapor, the resulting dry air would have a very constant gaseous composition. This makes it possible to regard atmospheric air (for meteorological purposes) as a mixture of dry air and water vapor. Below we give the principal physical characteristics of dry and actual moist (vapor-containing) air in the lower regions of the atmosphere. The composition of the upper, highly rarefied layers of the atmosphere is considered separately.

§ 1. Composition of dry air in the lower atmosphere

The composition of dry air remains constant with regard to its principal constituents up to considerable heights. Table 3 cites data currently accepted for the lower layers of the atmosphere (up to an altitude of 20-25 km). From this table it is evident that over 99.0% of dry air consists of molecular nitrogen and oxygen; aside from the argon and carbon dioxide always present in amounts slightly below 1%, all remaining gaseous traces account for less than 0.005% by volume. The molecular weight of dry air is $28.966 \approx 29$, very close to the molecular weight of the predominant gases nitrogen and oxygen.

We might mention that a characteristic feature of the earth's atmosphere is its extremely low content of light (H_2, He, CH_4) and heavy (Kr, Xe) gases.

By way of supplement to Table 3, we give certain basic molecular characteristics of dry air in Table 4. From here it is seen that dry air is about 800 times lighter than water, and one cubic meter of air at normal temperature and pressure ($t = 0°$ and $p = 760$ mm Hg) weighs $1.293 \text{ kg} \approx 1.3 \text{ kg}$. Its heat capacity at constant pressure is roughly four times lower than that of water, and its calorimetric conductivity is 20–30 times lower than that of water and thousands of times lower than that of metals; the thermometric

conductivity of air, on the other hand, is fairly high, higher than that of soil or metals.

TABLE 3

Composition of dry air

Gases	Molecular weight	Content (percents by volume)	Density		Critical temperature* t°
			absolute (g/m^3) at 760 mm, 0°	relative to dry air	
Nitrogen	28.016	78.084 ± 0.004	1250	0.967	−147.2 (33.5)
Oxygen	32.000	20.946 ± 0.002	1429	1.105	−118.9 (49.7)
Argon	39.944	0.934 ± 0.001	1786	1.379	−122.0 (48.7)
Carbon dioxide	44.010	0.033 ± 0.001	1977	1.529	31.0 (73.0)
Neon	20.183	$(18.18 \pm 0.04)10^{-4}$	900	0.695	−228.0 (26.0)
Helium	4.003	$(5.24 \pm 0.004)10^{-4}$	178	0.138	−258 (2.3)
Methane	16.04	$\approx 2.2 \cdot 10^{-4}$	717	—	—
Krypton	83.7	$(1.14 \pm 0.01)10^{-4}$	3736	2.868	−63 (54.0)
Nitric oxide	44.016	$(0.5 \pm 0.1)10^{-4}$	1978	—	—
Hydrogen	2.016	$\approx 0.5 \cdot 10^{-4}$	90	0.070	−239 (12.8)
Xenon	131.3	$(0.087 \pm 0.001)10^{-4}$	5891	4.524	16.6 (58.2)
Ozone	48.000	Highly variable; $(0-0.07)10^{-4}$ near surface, $(1-3)10^{-4}$ at height of 20-30 km	2140	1.624	5 (92.3)
Dry air	28.966	100	1293	1.000	−140.7 (37.2)

* In the last column the number in brackets indicates the pressure (in atmospheres) corresponding to the given temperature.

The critical temperatures of all the atmospheric gases, with the exception of carbon dioxide and water vapor, are decidedly lower than the temperatures which prevail in the atmosphere. For carbon dioxide the critical temperature is higher, but its partial pressure in the atmosphere is low. All these factors ensure that the dry air is in the gaseous state, with properties which are very similar to those of an ideal gas.

We note that the molecular characteristics cited in Table 4 refer to immobile calm air. Under natural conditions where continuous mixing of the air takes place entirely different values of the viscosity and calorimetric and thermometric conductivity must be employed (this will be shown later in greater detail); these values are usually tens and hundreds of thousands of times greater than the molecular ones, and are dependent on the intensity of mixing in the atmosphere.

Let us dwell on the characteristics of a few of the gaseous constituents of the atmosphere.

No regular or remotely significant variation with time and geographic location can be discovered in the nitrogen, oxygen and argon contents. True, analyses of certain samples taken at considerable heights (over 18 km) through the agency of stratospheric balloons and automatic instruments have revealed a slight falling-off of the oxygen content with increasing altitude; however, the dispersion of the scant data obtained is fairly large, and therefore no reliable numerical estimate of the magnitude of this decrease

can be made. In any event, up to a height of about 25–30 km this decrease constitutes no more than a few tenths of a percent.

TABLE 4

Characteristics of dry air

Density $\begin{cases} \text{for } T = 273^\circ \text{ and } p = 760 \text{ mm Hg}, \rho_0 = 1.293 \cdot 10^{-3} \text{ g/cm}^3 \\ \text{for } T = 273^\circ \text{ and } p = 1000 \text{ mb}, \rho_0 = 1.276 \cdot 10^{-3} \text{ g/cm}^3 \end{cases}$

Volume of 1 g · mol. at $T = 273^\circ$ and $p = 760$ mm Hg, $v = 22.41$

Molecular weight $\mu = 28.966$

Heat capacity at constant pressure $c_p = 0.2388 = 0.24$ cal/g · deg

" " " " volume $c_v = 0.1712 = 0.17$ cal/g · deg

Ratio $\frac{c_p}{c_v} = \varkappa = 1\ 404 \approx 1\ 4$

Coefficient of calorimetric conductivity at $T = 273^\circ$; $\lambda = 5.8 \cdot 10^{-5}$ cal/cm · sec · deg

Coefficient of thermometric conductivity $a^2 = \frac{\lambda}{c_p \rho} = 0.17 \text{ cm}^2/\text{sec}$

Coefficient of thermal expansion (0–100°) $\alpha = 0.00367 = 0.004 \text{ deg}^{-1}$

Viscosity $\eta = 1.72 \cdot 10^{-4}$ g/cm · sec

Slight variations are found in the carbon dioxide content, which is taken to be 0.033 % on the average. Despite its very low content the role of carbon dioxide in atmospheric phenomena is exceedingly great, for it is a strong absorber — and therefore emitter — of radiant energy in the region of long wavelengths. It is moreover necessary for the development of plant life. Carbon dioxide is delivered to the atmosphere by the earth's crust (volcanos and so on) and, since it is a product of the oxidation of all organic compounds, it is released during their combustion and oxidation as well as in the respiration of animal organisms. As a result the amount of carbon dioxide in the air varies slightly in time and with local conditions. For instance, its content is lower in daytime than at night; it varies seasonally as well, being greater in winter and smaller in summer and autumn. The carbon dioxide content also differs slightly from place to place. It is lower in Polar regions (about 0.02 % in Antarctica) and higher over land than over the ocean. The latter circumstance is due to the fact that carbon dioxide is absorbed from the air by sea water and subsequently deposited at the ocean bottom. This process is complicated by many factors (temperature, salinity of sea water, etc.) but there can be no doubt that the oceans act as a highly complex regulator of the carbon dioxide content.

The amount of carbon dioxide in the atmosphere also depends on purely local conditions; thus it is greater near volcanos and also near populated and especially industrial areas, where it sometimes reaches 0.05 % by volume.

The available data have made it possible to establish a continuous increase in the mean CO_2 content of the atmosphere over the last few decades; thus it was estimated to be 292 cm^3/m^3 before 1900; by 1935 it had risen to 310 cm^3/m^3; at present, having grown steadily, it is about 330 cm^3/m^3. This increase may perhaps be attributed to the expansion of industry.

The mean CO_2 content hardly changes with altitude, at least up to 20 km; only above this level does a slow decrease set in, judging by certain data.

As to the other gases present in the air in insignificant amounts, the most interesting are the lightest gases, hydrogen and helium, and also the allotropic compound of oxygen, ozone (O_3). The latter, in view of its greater significance, will be later considered in detail in § 7. Here we merely note that in the lower layers of the atmosphere (up to a height of about 10 km) the ozone content is low, varies in time, and changes irregularly with height; its volumetric content in these layers does not exceed $0.07 \cdot 10^{-4}$ %.

The average hydrogen content at the surface is commonly thought to be $5 \cdot 10^{-5}$ %; the variation of its content in the lower atmosphere and with altitude has scarcely been studied.

As to helium, released from the earth's surface as a product of the decay of radioactive substances, its content in the ground layers of the atmosphere is fairly constant; the deviations from the content indicated in Table 1 ($5.24 \cdot 10^{-4}$ %) are negligible. The helium content increases slightly with altitude; thus air samples taken in England at an altitude of 21–23 km indicated a content of $5.64 \cdot 10^{-4}$ % helium. A further slight increase of the helium content with elevation is evidenced by samples taken with rockets at altitudes of 41–72 km. Whether the helium content increases higher up as well, and how much helium occurs at very great heights, are questions which we will consider later.

Apart from those gases mentioned in Table 3, the atmosphere also contains — in perfectly insignificant and exceedingly variable amounts — certain other gases of natural and industrial origin as well as a whole series of chemical compounds such as SO_2, CO, H_2O_2 and the many nitrogen compounds (NO, N_2O, N_2O_4, N_2O_5). A very important fact is the presence, in the atmosphere, of radioactive emanations released by the earth's crust and their decay products, the radiations of which control, to a certain extent, the formation of ions in the atmosphere. The radon (Rn) content is negligible and can be taken to be $6 \cdot 10^{-18}$ % by volume for the entire atmosphere.

§ 2. Equation of state of dry air

It is well known that the principal parameters characterizing the state of a gas are: the temperature T, the pressure (buoyancy) p and the density ρ or specific volume $v = \frac{1}{\rho}$. The parameters p, T and $\rho = \frac{1}{v}$ are related by the so-called equation of state. For ideal gases this relation is expressed by the Mendeleev-Clapeyron formula

$$p = \frac{\rho}{\mu} R^* T, \tag{1}$$

where μ is the molecular weight of the gas and R^* the universal gas constant, numerically equal to the work (in ergs) done by one mole of gas upon heating by one degree at constant pressure. Numerically $R^* = 8.314 \cdot 10^7$ erg/mol · deg.

Let us return to equation (1) and consider it with reference to dry air. The molecular weight μ of such air is 28.966. Instead of the universal gas constant R^* it is convenient, as usual, to introduce the specific gas

constant R referred to one gram of air:

$$R = \frac{R^*}{\mu} = \frac{8.314 \cdot 10^7}{28.966} = 2.87 \cdot 10^6 \text{ erg/g} \cdot \text{deg}.$$

Introducing R, we write the equation of state of dry air (1) in the form

$$p = \rho RT, \text{ or } \rho = \frac{p}{RT}, \tag{2}$$

and, introducing the specific volume $v = \frac{1}{\rho}$, we obtain

$$pv = RT. \tag{3}$$

At normal pressure p_0 and temperature $T_0 = 273°\text{K}$ ($t = 0°\text{C}$) $\rho_0 = \frac{p_0}{RT_0}$, and therefore for $T = T_0(1+\alpha t)$

$$\rho = \rho_0 \frac{pT_0}{p_0 T} = \rho_0 \frac{p}{p_0} \frac{1}{1+\alpha t}, \tag{4}$$

where $\alpha = \frac{1}{273} = 0.00366$; the value of ρ_0, as indicated earlier, is $1.28 \cdot 10^{-3}$ g/cm^3 for $p = 1000$ mb.

From formula (4) it is obvious that the density of dry air is directly proportional to the pressure p and inversely proportional to the absolute temperature T. For $t = 20°$ and $p = 840$ mb the density of dry air is $1 \cdot 10^{-3}$ g/cm^3 and one cubic meter of it weighs one kilogram.

§ 3. Equation of state of water vapor. The correlation between the various characteristics of atmospheric humidity

Earlier it was mentioned that the critical temperature of water vapor is higher than the air temperatures encountered in many instances in the atmosphere; strictly speaking, therefore, it should not be classed among ideal gases. However, as experimental investigations have established, the differences between the physical properties of water vapor and those of an ideal gas are so slight that it can be regarded in practice as an ideal gas; accordingly one can write its equation of state in the form

$$e = \frac{R^*}{\mu_{vap}} T_{vap} \rho_{vap} \tag{5}$$

where e is the partial pressure of water vapor; the other symbols have the same meaning as before and the subscript "w" indicates that they refer to water vapor.

Since $\mu_{vap} = 18.016$, the specific gas constant R_{vap} for water vapor is given by

$$R_{vap} = \frac{R^*}{\mu_{vap}} = \frac{8.314 \cdot 10^7}{18.016} = 4.615 \cdot 10^6 \text{ erg/g} \cdot \text{deg},$$

further $R_{vap} = R \frac{\mu_a}{\mu_{vap}} = 1.608\, R \approx 1.6R$, where μ_a and R refer to dry air.

Taking the above into account, we will derive some relations between the various characteristics of atmospheric humidity given in § 7, Chapter 1.

From the relation (5) it is evident that absolute humidity as the amount of water vapor in grams per cubic meter of air is

$$a = \rho_{vap} \cdot 10^6 = \frac{e \cdot 10^3}{R_{vap} T} 10^6 = \frac{217e}{T} \text{ g/m}^3. \quad (6)$$

Expressing the temperature in the Celsius scale, we easily obtain

$$a = \frac{217e}{273(1+\alpha t)} = 0.795 \frac{e}{1+\alpha t} \text{ g/m}^3. \quad (7)$$

or, approximately,

$$a = 0.8 \frac{e}{1+\alpha t} \text{ g/m}^3. \quad (7')$$

Here e is always given in millibars.

If we express e in millimeters of mercury column,

$$a = 290 \frac{e}{T} = 1.06 \frac{e}{1+\alpha t} \text{ g/m}^3. \quad (8)$$

In this case the coefficient $\frac{1.06}{1+\alpha t}$ *is unity for* $t = 16.4°$ and, therefore, $a = e$.

Owing to the closeness of the numerical values of a and e, the vapor pressure e is sometimes called the absolute humidity; this, of course, is incorrect.

Denoting the density of dry air by ρ_a, and expressing the mass of the dry air and the water vapor contained in one cubic meter in terms of ρ_a and ρ_{vap}, respectively, we write down the following expressions for the specific humidity q *and mixture ratio* s *in accordance with their definitions:*

$$q = \frac{\rho_{vap}}{\rho_{vap} + \rho_a} \quad \text{and} \quad s = \frac{\rho_{vap}}{\rho_a},$$

from which

$$q = \frac{s}{s+1} \quad \text{or} \quad s = \frac{1}{1-q}. \quad (9)$$

But

$$\rho_{vap} = \frac{e}{R_{vap} T} \quad \text{and} \quad \rho_a = \frac{p-e}{RT}, \quad (10)$$

where $(p-e)$ is the partial pressure of dry air, and therefore

$$q = \frac{\frac{e}{R_{vap}T}}{\frac{e}{R_{vap}T} + \frac{p-e}{RT}} \quad \text{and} \quad s = \frac{\frac{e}{R_{vap}T}}{\frac{p-e}{RT}}. \quad (11)$$

Bearing in mind that $R_{vap}\ 1.608R$, we obtain

$$\left.\begin{aligned} q &= \frac{e}{1.608\left[p - e + \frac{e}{1.608}\right]} = \frac{0.622e}{p - 0.378e} \text{ g/g} \\ s &= \frac{e}{1.608(p-e)} = 0.622 \frac{e}{p-e} \end{aligned}\right\} \quad (12)$$

Since in nature $p \gg e$ (and all the more so $p \gg 0.378e$), at all times, instead of (12) one can assume approximately that

$$q = s = 0.622 \frac{e}{p} \text{ g/g} = 622 \frac{e}{p} \text{ g/kg}. \quad (13)$$

In reality, as is clear from (12), s is always slightly larger than q; however, this difference is very small.

§ 4. Equation of state of moist air. Virtual temperature

Taking into account the foregoing section, we will now derive the equation of state for real (moist) air as a mixture of dry air and water vapor.

If air at a temperature T and pressure p contains water vapor the partial pressure of which is e, and if, therefore, the partial pressure of the dry air is $p-e$, then, from the equation of state, we obtain the equation

$$\rho_a = \frac{p-e}{RT} = \frac{p-e}{T}\frac{\mu_a}{R^*},$$

for the density of dry air ρ_a and the equation

$$\rho_{vap} = \frac{e}{R_{vap}T} = \frac{e}{T}\frac{\mu_{vap}}{R^*},$$

for the density of water vapor ρ_w, where μ_a and μ_w are the molecular weights of dry air and water vapor respectively.

Obviously, the density ρ of moist air is

$$\rho = \rho_a + \rho_{vap}$$

or

$$\rho = \frac{p-e}{RT} + \frac{e}{R_{vap}T}.$$

Since $R_{vap} = 1.608R \approx 1.6R$,

$$\rho = \frac{p-e}{RT} + \frac{e}{1.6RT} = \frac{p}{RT}\left[1 - \frac{e}{p} + \frac{e}{1.6p}\right] = \frac{p}{RT}\left[1 - 0.378\frac{e}{p}\right]. \tag{14}$$

As $e \ll p$, one can write down approximately

$$\rho \approx \frac{p}{RT\left[1 + 0.378\frac{e}{p}\right]} \tag{14'}$$

or, recalling that $q = 0.622\frac{e}{p}$,

$$\rho = \frac{p}{RT[1 + 0.608q]}.$$

The above is the equation of state of moist air.

If one introduces a certain arbitrary temperature T_v into the preceding formula by the relation

$$T_a = T\left[1 + 0.378\frac{e}{p}\right] = T[1 + 0.608q], \tag{15}$$

it takes on the form

$$\rho = \frac{p}{RT_v}, \tag{16}$$

analogous to expression (2).

The arbitrary temperature T_v has been named virtual temperature. The introduction of this quantity makes it possible to retain the gas constant of dry air in the equation of state for moist air; therein lies its great practical significance. The physical meaning of the virtual temperature can be clarified in the following way: it is the temperature at which the density of dry air is equal to the density of moist air at the same pressure.

It follows from (15) that the virtual temperature is always higher than the actual temperature of air, and therefore for the same p the density of moist air is always lower than the density of dry air. Table 5 gives the values of the so-called virtual temperature difference $T_v - T$ for saturated water vapor, for which $e = E$. We can find this difference from (15), taking (13) into account; indeed,

$$T_v - T = 0.378 \frac{E}{p} T = 0.608\, qT \approx 0.6\, qT. \qquad (17)$$

where q is the specific humidity.

TABLE 5

Virtual temperature differences for saturated water vapor

p (mb)	Temperature (deg)								
	-40	-30	-20	-10	0	10	20	30	40
1000	0.0	0.0	0.1	0.3	0.6	1.3	2.6	4.9	8.8
500	0.0	1.0	0.2	0.3	1.2	—	—	—	—

As is obvious from the table, the virtual correction can reach several degrees in extreme cases at high temperatures, but generally it does not exceed 1-2° near the ground, while at great altitudes (low temperatures) it is only a small fraction of a degree. Here it should be borne in mind that the difference $T_v - T$ decreases with decreasing humidity, which also falls off with altitude. Since this difference does not exceed the limits of accuracy of aerological observations for altitudes greater than 3-4 km, the actual temperature T can be used instead of T_v in this region. Correspondingly, the difference between the densities of dry and moist air is

TABLE 6

Density of dry and saturated air at a pressure of 1000 mb

Density of air (10^{-3} g/cm^3)	Temperature (deg)					
	-20	-10	0	10	20	30
Dry	1.38	1.32	1.276	1.231	1.189	1.150
Moist, saturated	1.37	1.32	1.273	1.225	1.178	1.132

insignificant at low temperatures but should be taken into account at high temperatures and humidities. The values of the density for various temperatures are given in Table 6.

§ 5. Water vapor in the atmosphere

Water vapor enters the atmosphere owing to the evaporation of water from the earth's surface and spreads through the atmosphere by mixing. At the same time the maximum content of water vapor is restricted by its saturation pressure; condensation of the vapor takes place when this point is reached.

The processes of evaporation and condensation are accompanied by the absorption or generation of heat, which is reflected in the thermal regime of the corresponding layers of the atmosphere. Water vapor also plays a large part in the processes of absorption and emission of radiation. All this explains the tremendous role of water vapor in all phenomena in the atmosphere; this question, together with the distribution of water vapor in the atmosphere, will be considered in detail in Part Four. Here we note that the mean vapor pressure e drops off rapidly according to the law

$$e = e_0 \cdot 10^{-\frac{z}{\beta}}, \tag{18}$$

where e and e_0 are the mean vapor pressure at the height z and at the surface respectively. The coefficient β is found from observational data. For the lower atmosphere β is about 5000 m (if z is expressed in meters). From this it follows that e decreases by a factor of 10 at a height of 5 km, 2 at 1.5–2 km, i.e., the vertical decrease of the vapor pressure is much faster than that of the overall pressure of the atmosphere.

The concentration of water vapor (volumetric, g/cm^3, or weight, g/g) also decreases (exponentially) with altitude.

Thus the vapor content falls off very rapidly with altitude and becomes negligible when a level of about 8–10 km is reached; at great altitudes the air is, as a rule, very dry. However, this does not exclude the possibility of vapor accumulating at certain very high altitudes, where under specific conditions it may condense and form thin clouds. For example, the so-called mother-of-pearl clouds are sometimes recorded at an elevation of 25–30 km. It may be as some researchers believe that the yet higher noctilucent clouds (altitude 80–85 km) have the same origin. Knowing the law of variation of a_z with altitude, one can express the total amount of water vapor W contained in a column of atmosphere of unit cross section (cm^2) up to any altitude z, which may even be the upper boundary of atmosphere, as the integral

$$W = \int_0^{\infty} a_z dz = a_0 \int_0^{z} e^{-\frac{z}{\beta}} dz. \tag{19}$$

But the value of the integral $\int_0^{\infty} e^{-\frac{z}{\beta}} dz$ gives the height of the so-called homogeneous atmosphere of water vapor and is numerically equal to $H_{vap} = 2080$ m, and therefore

$$W = H_{vap} a_0 = 0.208\, a_0 \text{ g/cm}^2, \tag{20}$$

where a_0 is expressed in g/m^3.

Let us assume that all the water vapor contained in a vertical column of the atmosphere one square centimeter in cross section condenses; the layer of water formed as a result of this condensation (if one expresses its

height in centimeters) gives us the total mass of water vapor contained in the given column of air. The height of the layer of settled water is frequently used as a convenient characteristic when evaluating the overall content of water vapor in the atmosphere.

§ 6. Ozone in the atmosphere

Among the gases contained in the atmosphere a very important one is ozone (O_3), present in a layer extending from the ground to a height of 55–60 km.

Although the total amount of ozone is limited its significance for atmospheric processes and life on earth is very great. This is because ozone is a strong absorber of solar radiation, especially in the region of ultraviolet rays with wavelengths shorter than $\lambda = 2900$ Å; as a result these biologically very active rays fail to reach the earth's surface, and the solar spectrum near the ground is sharply cut off at a wavelength slightly lower than 300 mμ (~285–290 mμ).

Apart from the ultraviolet, ozone also has absorption bands in the visible and infrared regions of the spectrum, a question which will be treated in detail in Chapter 8, § 3.

As a result of the absorption of solar energy in the upper part of the ozone layer the atmospheric temperature at these altitudes is considerably higher than it would be otherwise and the ozone layer acts as a virtual reservoir of heat in the atmosphere.

The ozone content of the lower layers of the atmosphere (to a height of about 10 km) is negligible. Its concentration near the surface, measured by chemical methods, averages $3 \cdot 10^{-5}$ g/m^3; however, it varies over a significant range (zero to 10^{-4} g/m^3) depending on the latitude, season and a variety of local factors.

As one moves upward from the earth's surface, the ozone content changes irregularly up to a height of 10 km, then begins to increase; this increase is particularly marked above 12–15 km. The maximum of the ozone content occurs at a height of 20–25 km; higher up the amount of ozone falls off gradually, becoming negligible at a height of 55–60 km.*

Data on the ozone content of the higher layers are obtained by analysis of air samples taken at these altitudes and (chiefly) by special optical methods involving measurements of the intensity of the absorption bands of ozone. Valuable data have been obtained in the course of stratosphere balloon ascents and also, very recently, by means of rockets carrying spectrographs for recording the solar spectrum. The vertical variation of the amount of ozone can be determined from the intensity of absorption in the ultraviolet region of the spectrum. Figure 1 gives an example of the variation of the solar spectrum, photographed during the ascent of a V–2 rocket (14 June 1949).

It is customary to express the total ozone content of the atmosphere in terms of the thickness (in centimeters) of the layer which would form if all the ozone contained in an atmospheric column of unit cross section were brought to normal pressure and temperature (760 mm Hg and 273° K). This

* We note that if the relative volumetric content of ozone (ozone/air volume ratio) is considered the maximum of this ratio occurs at a somewhat higher level, namely 30–35 km.

is the so-called reduced thickness of the ozone layer. It averages about 3.0 mm for the entire atmosphere. Similarly, the amount of ozone in the various layers is usually expressed in terms of the thickness of the reduced layer of ozone (in centimeters) as determined for a column of unit cross section one kilometer high.

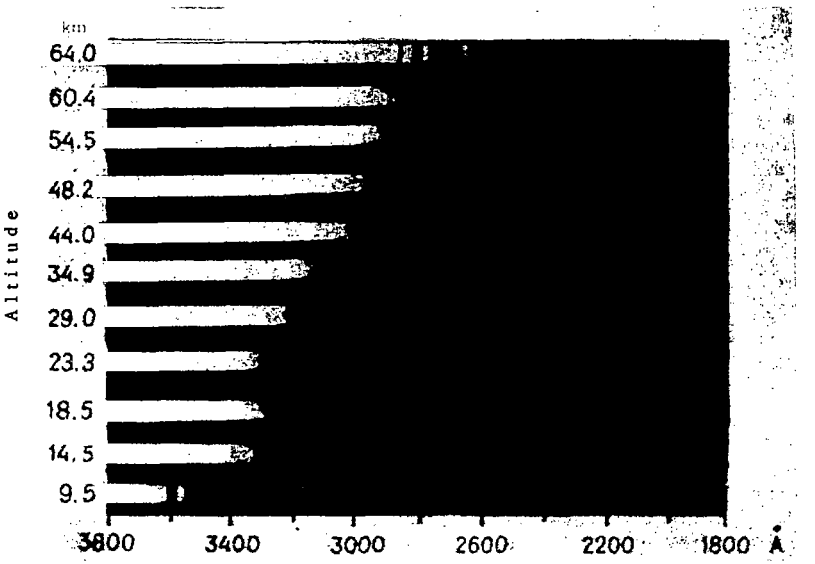


FIGURE 1. Solar spectrum obtained during ascent of V-2 rocket, 14 June 1949*

The vertical distribution of ozone varies somewhat from point to point and is subject to certain variations in time. These variations are particularly marked in the lower part of the ozone layer (10–25 km). Here the fluctuations of the ozone content are roughly three times greater than the variations of the overall ozone concentration over the entire thickness of the atmosphere; these fluctuations, moreover, increase with the total ozone content. In many cases the position of the concentration maximum is different from the mean indicated position, and sometimes (for large ozone contents) a secondary maximum is observed at altitudes below 20 km.

Examples of the vertical distribution of ozone are shown in Figure 2.

The question of the formation of the ozone layer (ozonosphere) and the characteristic features of its distribution is resolved to some extent by the photochemical theory, founded by Chapman and developed further by a number of investigators. According to this theory, the primary reaction responsible for the formation of ozone is the dissociation of the oxygen molecule O_2 during the absorption of solar ultraviolet in the wavelength range 1300–1750 Å; this can be expressed as $O_2+h\upsilon \rightarrow O+O$, where $h\upsilon$ is the energy of the active radiation, h the Planck constant and υ the frequency.

* [The poor quality of this reproduction is due to the state of the original Russian illustration.]

In the presence of a third neutral molecule, M, the oxygen atoms formed in the collisions combine with an O_2 molecule to form ozone molecules in the reaction $O_2+O+M\rightarrow O_3+M$. In addition to this reaction, other reactions leading to the formation of O_3 are also possible; thus

$$O_2+O_2\rightarrow O_2+O+O=O_3+O. \qquad (21)$$

Simultaneously with the formation of O_3, its destruction by photodissociation and in collisions with oxygen atoms also takes place. Thus the inverse reaction $O_3+h\nu\rightarrow O_2+O^*$ (O^* being an excited oxygen atom) takes place when ozone absorbs solar radiation in definite wavelengths, and a reaction of the type $O_3+O\rightarrow O_2+O_2$ takes place in collisions.

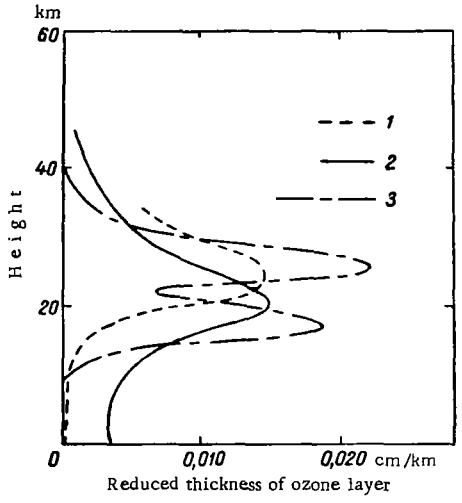


FIGURE 2. Vertical distribution of ozone

1-from data of sounding balloon released near Washington, 1938 (after Koblents and Stair); 2-from calculations based on optical measurements, 1935-1942 (after Meetham and Dobson); 3-from data of V-2 rocket, White Sands, 1946, (after Newell and Siry).

As these basic processes of production and destruction operate simultaneously, an equilibrium state should become established. However, the intensity of the processes varies slightly with time and place and as a result the imminent equilibrium may constantly be disrupted; restoration of the equilibrium, moreover, is very slow. This leads to variations in the ozone content and its height distribution. The processes of horizontal and vertical mixing which take place in the atmosphere may lead to the transfer of ozone from one region to another, and moreover over considerable distances. To a significant extent, this may account for the large and nonperiodic changes observed at altitudes of 10 to 20 km.

The ozone content in the lower layers of the atmosphere (up to a height of 10 km) is influenced by such factors as the decay of ozone in collisions with cloud particles, and also local factors leading to the formation of

ozone, e.g., lightning discharges or the oxidation of certain organic substances.

The total amount of ozone in the atmosphere is subject to considerable changes in time and varies also with latitude; the observed extrema are contained roughly within the range 1.5–5.5 mm. It has been established that the amount of ozone in the atmosphere is minimal over the equator; the variations of its content over the year are also small here. The ozone content increases in the direction of the high latitudes and reaches maximal values in the polar regions where strong seasonal fluctuations characterized by a maximum in the spring months and minimum in autumn are also observed.

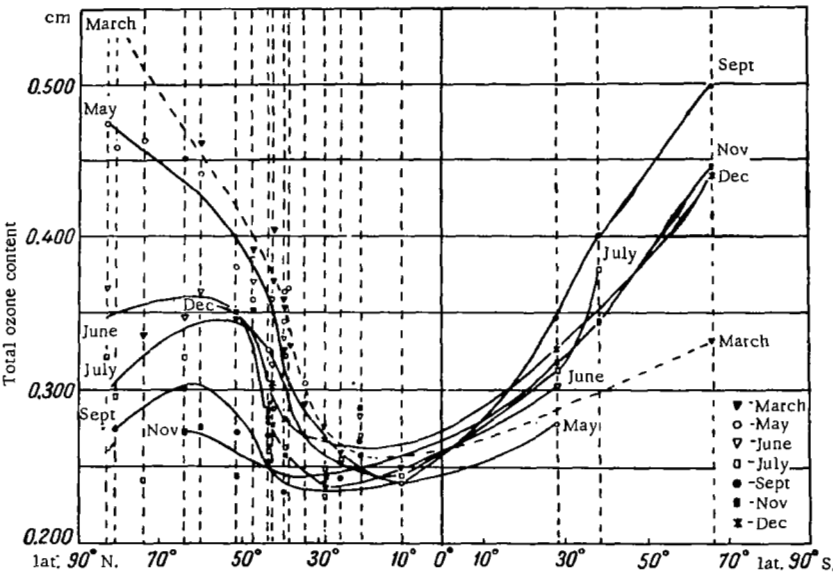


FIGURE 3. Latitudinal variation of the overall ozone content of the atmosphere in different months, according to data of observations during the IGY period (after G. P. Gushchin)

This pattern of annual variation is maintained steadily even in individual years, a point confirmed by data obtained during the IGY period by Soviet ozonometric stations (Figure 3). Daily fluctuations of the overall ozone content are limited, since there is no great difference between the mean nighttime and daytime values.

The nonperiodic fluctuations are of particular interest. Thus the mean amount of ozone varies somewhat from year to year; the available observational material, however, does not as yet allow one to establish any clear-cut regularity in these variations (e.g., a connection with the eleven-year cycle of solar activity, a possibility mentioned by certain authors).

Day-to-day variations — especially large in spring — are also observed, as well as nonperiodic variations connected with changes in weather conditions; the latter are of particularly great interest for meteorologists. Thus it has been established that over certain regions of the earth's

surface the ozone content rises when high-latitude air flows in at heights of about 16 km and drops when air moves in from tropical latitudes. A connection has also been observed with changes in pressure (to a greater pressure at the surface corresponds a lower ozone content), temperature and other meteorological elements. Exceedingly rapid and moreover sharp variations in the ozone content which create the impression that an "ozone cloud" is passing overhead are also observed occasionally.

The connection between the ozone content and vertical motions in the atmosphere is indisputable.

The nonperiodic fluctuations connected with changes in the state of the weather clearly indicate the possibility of ozone transport by air streams in the lower regions of the ozone layer. The peculiar character of the vertical distribution of ozone also tends to support such a possibility.

§ 7. Atmospheric aerosols

Minute particles of various substances in the solid and liquid state, of both natural and industrial origin, are always present in suspension in real atmospheric air. These particles vary widely in chemical composition, size, form and physical properties. Some may be electrically charged; among these should be classed the so-called ions (cf., Chapter 29).

Owing to the small size of the particles (their radius for the most part does not exceed $10-20\mu$) they remain in a suspended state in the atmosphere for a long time and are transported by vertical and horizontal streams of air. The collective term "aerosol" is applied to the system of these dispersed particles. The atmosphere can be treated by analogy with the concepts of physical chemistry as a colloidal solution, in which, however, the solvent is not a liquid but a gaseous medium (the air) and the dissolved substance suspended in it is the above-named minute admixtures.

Side by side with these particles, under specific conditions larger particles of moisture (droplets of precipitation and ice crystals) as well as large particles of dust are also found in the atmosphere; the latter in view of their large mass cannot remain long in the atmosphere and settle down on the surface fairly rapidly.

All these particles are of very great significance for the development of a number of atmospheric phenomena. For example, a significant fraction of these act as so-called condensation nuclei (Chapter 19, § 4); the condensation of water vapor in the atmosphere begins on such nuclei, and their presence is absolutely necessary for the formation of fog, clouds and, finally, the elements of precipitation (raindrops, snowflakes, etc.). Furthermore, their presence determines the degree of turbidity of the atmosphere and has a strong influence on the flux of solar radiation, as well as being the cause of a number of optical phenomena. Many phenomena of atmsopheric electricity are due to charged aerosols (ions). Finally, the dust content of the atmosphere is an important hygienic factor, especially near industrial towns. All this explains the wide interest aroused by the study of aerosols and dust, which is indeed being carried out by a wide variety of different methods.

Aerosols can be classed according to origin into two large groups: 1) particles of terrestrial origin and 2) particles of cosmic origin (cosmic

dust), which enter the atmosphere from cosmic space and are formed inside it during the vaporization of meteors.

The first group includes:

a) particles of soil and products of rock weathering,

b) particles of organic origin (microorganisms, plant pollen, etc.),

c) particles of smoke of industrial origin as well as those formed in forest and peat fires,

d) volcanic dust ejected by active volcanos, occasionally rising to considerable altitudes (up to 30 km),

e) sea-salt particles (maritime nuclei) entering the atmosphere in the evaporation of sea spray,

f) particles of industrial origin, more specifically products of the operation of industrial enterprises.

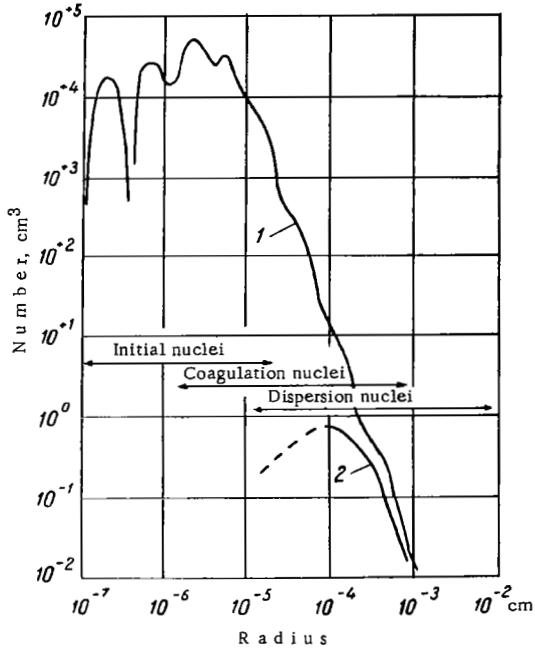


FIGURE 4. Size distribution of condensation nuclei (after Junge)

1–Aitken nuclei, large and giant; 2–sea-salt particles.

Particular attention has been paid in recent years to the study of aerosol particles which enter the atmosphere in large amounts during the testing of atomic weapons. Owing to the presence in them of radioactive substances they are of wide biological significance.

Figure 4 shows the average distribution of aerosol particles according to size, based on the experimental researches of Junge. The maximum of the number of particles coincides with a size of the order of $10^{-6}-10^{-5}$ cm.

The complete spectrum of aerosol particles, as is evident from Figure 4, ranges up to $r \approx 10^{-3}$ cm.

The amount of dust and condensation nuclei contained in the air near the surface of the ground in different parts of the world is highly variable. Naturally, the concentration of aerosol particles is greater near the sources which transfer them into the atmosphere and is usually small in pure country, mountain or sea air. The highest concentrations of particles are recorded in cities and especially in industrial cities, where their number in one cubic centimeter can be in the thousands or even hundreds of thousands. The degree of pollution of the air over large towns is seen from tentative calculations according to which the air over industrial cities contains tens of thousands of tons of soot and dust.

Table 7 gives the average values of the concentration of condensation nuclei under various conditions (after Simpson).

TABLE 7

Concentration of condensation nuclei

Locality	Number of nuclei per cm^3		
	average	maximum	minimum
Urban	150 000	4 000 000	3500
Rural	30 000	400 000	600
Ocean	900	40 000	2

If small the aerosol particles are carried large distances away from their source by air currents and settle very slowly out of the atmosphere, in which they occur as if in suspension. The rate of fall of dust particles can be estimated tentatively on the basis of the following elementary considerations.

Let us denote the mass of a dust particle by m and its radius by r (assuming that it has a spherical shape). Then in calm immobile air its rate of uniform fall is given by the relation

$$mg = f. \tag{22}$$

For small spherical particles the force of resistance of the medium f is given by the relation (Stokes' law)

$$f = 6\pi\eta r v, \tag{23}$$

where η is the coefficient of viscosity ($\eta = 1.72 \cdot 10^{-4}$ g/cm · sec at 0°) and v the rate of fall.*

Bearing in mind that the weight of the particle $mg = \left(\frac{4}{3}\pi r^3 \delta\right)$, where $\delta = \rho - \rho_a$, ρ is the density of the particle and ρ_a the density of the air, we find that

$$v = \frac{mg}{6\pi\eta r} = \frac{2}{9}\frac{g}{\eta} r^2 (\rho - \rho_a) \tag{24}$$

or, in view of the smallness of ρ_a compared with ρ, we obtain

$$v = \frac{2}{9}\frac{g}{\eta} \rho r^2. \tag{25}$$

* [Also "terminal velocity" in Western usage.]

In particular, assuming that $\rho = 1$, which holds, for example, for water droplets, we obtain for their rate of fall

$$v = 1.26 \cdot 10^6 r^2 \text{ cm/sec} \tag{26}$$

and, asssuming $r = 10^{-4}$ cm, we find that $v = 1.26 \cdot 10^{-2}$ cm/sec.

These formulas show how small the rate of fall must be for aerosols with dimensions in fractions of a micron.

The vertical distribution of aerosol particles is determined by a number of factors, the main ones being horizontal and vertical mixing in the atmosphere, the coagulation of particles, their settling under the influence of gravity and also their removal from the atmosphere as a result of the formation on them of cloud droplets and raindrops (or crystalline elements).

The law of vertical distribution of aerosols can be found theoretically under certain special assumptions. In the simplest case it is assumed that for a steady (average) distribution the upward stream of particles due to turbulent mixing compensates the stream of settling particles. This condition can be written down in the form

$$vn = -k \frac{dn}{dz}. \tag{27}$$

Here n is the concentration of particles of a definite size (i.e., their number per cubic centimeter), v their rate of fall and k the coefficient of turbulence* (cm^2/sec) characterizing the intensity of mixing.

Integrating this equation from $z = 0$, where $n = n_0$, to z, where $n = n_z$, for constant k, we obtain

$$n_z = n_0 e^{-\frac{v}{k} z}. \tag{28}$$

It follows from this relation that for more intensive mixing (large values of k) the dust content decreases more slowly with altitude and, therefore, is distributed over a greater height. It is also obvious that the content of larger and heavier particles (greater velocity v) decreases more rapidly with altitude than the content of finer and lighter particles. However, the above relation is largely unsuitable for practical calculations, primarily because particles are of various sizes but also because there are other factors which can modify their number in the volume under consideration. In particular, the process of coagulation may be of considerable importance for large particle concentrations.

E. S. Selezneva and M. I. Yudin solved the problem of the vertical distribution of aerosols with an allowance for particle coagulation. In view of the fact that for the overwhelming number of particles v is very small ($10^{-4} - 10^{-5}$ cm/sec), the settling of particles can be disregarded. The equation of transport of aerosols can then be written in the form

$$k \frac{d^2 n}{dz^2} - \alpha n^2 = 0, \tag{29}$$

where α is the so-called constant of coagulation and the term αn^2 accounts for the process of coagulation in the aerosol.

After integrating (29) we obtain the formula

$$n_z = n_0 \frac{C^2}{(z + C)^2}. \tag{30}$$

* [See Ch. III, § 2.]

The integration constant $C = \sqrt{\frac{6k}{an_0}}$ can easily be found from observational data from the condition

$$C = z \quad \text{for} \quad \frac{n_z}{n_0} = \frac{1}{4}.$$

Formula (30) has been verified against the large volume of material on condensation nuclei obtained in observations carried out during the IGY period in the Soviet Union. As is evident from Figure 5, it provides a good description of the average distribution of condensation nuclei up to a height of 5–6 km. It is also seen from this figure that the exponential law holds

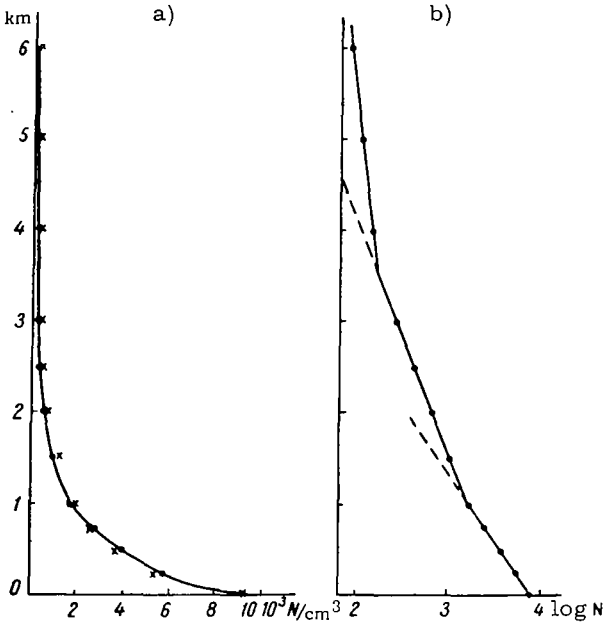


FIGURE 5. Average vertical distribution of condensation nuclei over the European USSR, from data of observations during the IGY period (after E. S. Selezneva)

a–variation of nucleus concentration N with height; b–variation of log N. Dots indicate experimental data, crosses data calculated according to formula (30).

only in those layers of the atmosphere in which the following linear relation is valid

$$\lg n = \lg n_0 - az.$$

The distribution in the atmosphere of the large particles of aerosol ($r > 1\mu$) has been studied abroad by Woodcock, Moore, Mason and others in connection with the problem of formation of precipitation. In recent years thorough research on such particles has been carried out by A.G. Laktionov with the aid of a specially designed optical instrument. These researches

have established that the concentration of large particles in the free atmosphere is several tens per liter of air.

In conformity with general theoretical ideas, the observations indicate large changes in the dust content at various heights during the day. At night with the attenuation of mixing the particles become concentrated in the lower layers, i.e., their settling takes place, while in daytime, on the contrary, they are disseminated over greater altitudes. In general, however, the character of the daily variation is complex, as it depends not only on the general cause mentioned here but also on local conditions which control the entry of dust into the atmosphere; in mountain conditions, for example, the daily pattern may be the exact opposite of that described above. Similarly the annual variation of the dust content is also complex. In populated and especially industrial areas the maximum dust content in the ground layer of air frequently occurs in the winter months and the minimum in the summer months; away from industrial areas the opposite pattern is observed.

In conclusion we will say a few words concerning cosmic dust in the atmosphere. The separate investigation of cosmic dust is hampered by considerable difficulties. Indirect considerations lead to the conclusion that the total amount entering the earth's atmosphere is very considerable and may roughly be estimated at many tens or even several hundreds of tons (up to 1000 tons) daily. Cosmic dust enters the atmosphere irregularly during the year and penetrates very slowly from the upper layers to the surface, where it settles. Its concentration and vertical distribution are not known exactly, but since a significant fraction of the total is formed in the vaporization of meteors it might be supposed that the maximum concentration occurs at an altitude of the order of 70–90 km. Cosmic dust doubtless plays a certain part in the thermal regime of the upper atmospheric layers.

§ 8. Composition of air at high altitudes

For a long time efforts were made to obtain some idea of the composition of atmospheric air at high altitudes from Dalton's law, according to which the total pressure p of a mixture of gases is equal to the sum of the partial pressures of the individual gases p_i, i.e.,

$$p = \sum_{i=1}^{n} p_i.$$

It was assumed that atmospheric air was immobile and not subject to any external influence other than that of gravity. Given this assumption, owing to the process of heat conduction isothermy should be established throughout the atmosphere; that is, the same temperature $T = \text{const}$ should prevail throughout. In an isothermal atmosphere the pressure p_i and density ρ_i of any gas will vary with height according to the law (cf. Chapter 5):

$$(p_i)_h = (p_i)_0 e^{-\frac{g\mu_i}{R^0 T}h} = (p_i)_0 e^{-\frac{h}{H_i}},$$

$$(\rho_i)_h = (\rho_i)_0 e^{-\frac{g\mu_i}{R^0 T}h} = (\rho_i)_0 e^{-\frac{h}{H_i}},$$

where $(p_i)_0$ and $(\rho_i)_0$ are the values of p_i and ρ_i at the initial level, g is

the acceleration of gravity (assumed to be constant), μ_i the molecular weight of the gas, R^* the universal gas constant, T the absolute temperature, h the height and H_i the so-called scale height or height of the homogeneous atmosphere for the gas under consideration at the temperature T (for the values of H see Table 13).

Given this premise the relative content of lighter gases (with smaller values of μ_i), tne partial pressures of which drop off more slowly, should rise with increasing altitude. Using the observed values of $(p_i)_0$ and $(\rho_i)_0$ at some initial level as initial data, one should be able, it would seem, to calculate the composition of the air at various altitudes as well. This constitutes the so-called hypothesis of the separation of gases by diffusion. Since observational data indicated constancy of the gaseous composition of air up to a height of at least 20 km it was assumed that the distribution of gases by diffusion was hampered up to this height by the presence of mixing, and that separation by diffusion began above this level.

The many investigators who performed the corresponding calculations using various initial data invariably obtained a prevalence of light gases in the air at high altitudes; moreover, it usually turned out that the earth's atmosphere should have consisted of helium and hydrogen starting from a height of about 100 km and entirely of hydrogen at a still higher level. However, this contradicts all data of investigations into the spectrum of aurora polaris, nightglow and other optical phenomena observed in the upper atmosphere, and also fails to be confirmed by other data obtained in the last decade concerning the composition of the atmosphere at high altitudes. Thus the "hypothesis of light gases in the atmosphere" has now been abandoned completely and remains only of historical interest.

Nonetheless the question of the separation of gases in the upper layers of the atmosphere and of the role of molecular diffusion in the processes which take place at these heights, remains an important one. It is certain that winds and turbulent mixing are present up to very great altitudes and that this hampers the separation of gases by diffusion; however, owing to the increase of the mean free path of the molecules with height the role of molecular diffusion also increases. But since the two processes — diffusion and mixing — act simultaneously, an important question arises: at what height does the tendency to establish diffusive equilibrium begin to predominate over mixing and how might this affect the variation of the composition of air with altitude? Estimates performed by a number of authors lead to a definite conclusion, namely that the lowest limit for the diffusion level (level at which diffusion may begin to predominate over the mixing process) should be close to 150 km. In the opinion of certain other authors, however, the level of separation of gases is significantly higher (between 300 and 400 km). Rocket studies in the USSR and other countries confirm that there is no basis for speaking of separation by diffusion before the 100 km level. It can be assumed that separation by diffusion begins to manifest itself starting roughly from a height of 200–300 km.

The composition of the atmosphere at high altitudes is determined from observations of the spectrum of aurora polaris, nightglow and twilight glow (Chapter 33), and the absorption of solar radiation; in addition rocket research involves the use of radio-frequency mass-spectrometers and the analysis of samples of air from various altitudes. On the basis of the data obtained it has been established that the principal gases in the atmosphere

up to the highest levels accessible to research by spectroscopic methods (about 1000 km) are nitrogen and oxygen. At a height of 90–95 km the relative content of these gases differs only slightly from their content at the surface (roughly 20% O_2 and 80% N_2 by volume). Above 90 km a marked dissociation of the oxygen molecules makes its appearance; this dissociation increases very rapidly with altitude, and at a height of about 130 km the content of molecular oxygen is only about 25%.

At high altitudes nitrogen is also partially in the atomic state. However, the mechanism of dissociation of nitrogen is far more complex than that of oxygen and is still incompletely clear. Information relating to the distribution and relative content of atomic nitrogen is more restricted. Nevertheless there exists an entire series of reliable proofs that the dissociation of nitrogen begins at a higher level than that of oxygen (about 220 km) and that the degree of its dissociation increases comparatively slowly with height. According to observations with a radio-frequency mass-spectrometer mounted on the third Soviet satellite, between a height of 230 and 820 km the degree of nitrogen dissociation increased only by a factor of (roughly) two. Thus one can assume, as previously suggested on the basis of spectroscopic investigations, that molecular nitrogen is still present in the region of the highest aurora polaris (altitude roughly 1000 km). True, the number of atoms of nitrogen present here is already considerably larger than the number of molecules.

The dissociation of oxygen and nitrogen raises the important question of the molecular weight of the air at heights exceeding 100 km. Here one should allow for the fact that aside from dissociation the average molecular weight of air is also influenced by its dynamic state (i.e., mixing or the separation of gases by diffusion, since in the presence of the latter the molecular weight ought to diminish with altitude). This complex problem is solved variously by different investigators. However, it has been established that the molecular weight of air to a height of the order of 90 km remains unchanged and is 28.97 g/mole; above this level it decreases due to dissociation of O_2 to 23.95 g/mole at about 130 km. Still higher the molecular weight of air continues to decrease as a result of nitrogen dissociation and partly of diffusion. Where the subsequent decrease ends is still too early to say.

In addition to oxygen and nitrogen, it has been established that a number of other gases are also present in small quantities in the upper layers. The question of the content of the lightest gases — helium (He) and hydrogen (H_2) — in the upper atmosphere is a particularly captivating one for researchers.

With regard to helium, detected in the lower layers, as indicated earlier, up to a height of several tens of kilometers, spectroscopic data fail to give any indication of its presence in perceptible amounts at heights above 100 km. The absence of emission lines of helium could be a consequence of the fact that the proper conditions for their appearance are lacking at these altitudes. At the same time, estimates of the total helium content in a vertical column of air of unit cross section based on data pertaining to its content at the surface lead to a value which is many times smaller than the amount released from the crust in the earth's elapsed life. This implies that the helium continually released by the earth's crust does not remain in the atmosphere. According to tentative calculations, about

10^5 helium atoms should be leaving the outer layers of the atmosphere per second per square centimeter. A similar escape (dissipation) into outer space — indeed a more intensive one — obtains in the case of hydrogen. However, its lines, although weak, are nonetheless observed occasionally in the spectra of upper-air auroras. Certain indirect signs indicate the presence of a thin layer of molecular hydrogen (H_2) with a slightly increased concentration (up to about 10^{10} cm^{-3}) at a height of about 70 km; but at greater elevations (> 100 km) the hydrogen is indisputably dissociated.

The shift of the hydrogen lines ($\lambda = 4860.7$ Å and $\lambda = 6563$ Å) toward the ultraviolet end of the spectrum in the spectra of certain auroras points to the presence in the upper layers of the atmosphere of hydrogen atoms moving at very high speeds (roughly up to 1500 km/sec or more) toward the surface. The appearance of atoms with such speeds in the earth's atmosphere is possible as a result of incursions from the sun or from cosmic space.

Investigations into the question of the presence of hydrogen in the upper layers led to the detection of hydroxyl (OH) at altitudes of the order of 75–80 km; this compound is completely absent both from the lower layers and, apparently, from heights above 100 km. Its formation can be explained by various reactions and in particular by the collision of ozone molecules and hydrogen atoms.

The presence of OH is highly significant as it may explain the formation of molecules of water vapor (H_2O) at these heights. At the same time it is evidence of the existence in the atmosphere of free hydrogen, the presence of which gives rise to the possibility of reactions leading to the formation of various hydrogen compounds.

The spectra of nightglow and twilightglow were found to contain yellow sodium D lines (especially bright in twilightglow), proving the presence of sodium in the upper layers of the atmosphere. It was established that the sodium lines are emitted by two layers, one in the region between 35 and 60 km and the other at a height of about 250 km. However, it is highly probable that sodium is present in the entire layer from 35 to about 250 km; its relative concentration increases at greater altitudes.

The approximate overall sodium content is estimated to be 10^{10} atoms in a vertical column of atmosphere of unit cross section (1 cm^2). The nonuniform distribution of the intensity of the emission lines of sodium over the sky and the presence of seasonal fluctuations (maximum in winter and minimum in summer) point to an uneven distribution in the atmosphere and to variations in its content.

The question of whether the sodium in the upper layers comes from the earth's surface or interplanetary space has not yet been answered conclusively.

Certain observations of the absorption spectrum of the sun indicate the presence in the atmosphere of a number of isotopes, e.g., H^2, C^{13}, N^{15}, O^{17} and O^{18}.

Thus all the available data indicate that the earth's atmosphere is composed of nitrogen and oxygen throughout, as in the lower regions, and that the only distinction is that in the upper layers these gases are completely or partially dissociated and occur in the atomic state.

Chapter 3

TURBULENT MIXING IN THE ATMOSPHERE

§ 1. Turbulent character of motions in the atmosphere

The atmosphere is an extremely mobile medium in which motions differing very widely in scale, direction and velocity are perpetually taking place. Only in exceptional cases are these motions laminar (i.e., undisturbed, steady), when all particles move along smooth parallel trajectories. In such motions the interactions between individual particles take place exclusively owing to the thermal motion of molecules. Usually, however, all motions in the atmosphere are turbulent and characterized by a nonconstant velocity field. In turbulent motion the trajectories of individual air particles are highly irregular, wavy, and rapidly variable in time and space. Disorderly air streams, alternately strong and weak and variable in direction, are constantly being formed. Eddies of all sizes develop; elementary air masses arising in and separated from the general stream develop and move independently as integral formations for some time, then gradually disintegrate and are absorbed in the surrounding air. All this leads to intensive intermingling and interaction among the various parts of the medium.

The differences in the character of the motion can be illustrated by the well-known textbook experiment on the flow of a liquid in long straight pipes. A thin colored jet introduced into such a flow will retain its integrity and be broadened only slightly in the direction of flow as long as the liquid in the pipe moves slowly. As the velocity of the flow increases the colored jet first becomes wavy, then breaks down into several parts and is rapidly diffused over the entire stream. A motion of this kind is termed turbulent. It is largely rotational but the rotation exhibits no regularity and the path of each individual particle is highly complex and irregular, even though as a whole the liquid moves steadily in a definite direction. Turbulent motion has the following cause (as shown by experiments): when the velocity of a flow is increased above a certain critical point the viscosity of the liquid no longer suffices to damp out those perturbations of the velocity field which arise due to nonuniformity of the velocity or to the roughness of the boundary walls.

The turbulent character of air motions in the atmosphere can be studied by observing, for instance, the propagation of smoke issuing from a chimney. At low wind velocities (and temperature increasing with altitude) the smoke spreads over a large distance in the form of a thin jet. As wind speed increases the jet becomes wavy and sometimes (when temperature falls off rapidly with altitude) breaks down into several parts.

Turbulent motions in the atmosphere are clearly manifested in the phenomenon called wind gustiness. When one observes wind velocity and direction at a certain point with a low-inertia instrument one sees that both the velocity and the direction are somewhat different at every instant, deviating from the value averaged over a more or less considerable time interval.

Let us denote by u, v and w the instantaneous values of the components of the wind velocity along the coordinate axes at the instant t_0. The average values of these components over the interval T are

$$\overline{u}=\frac{1}{T}\int_{t_0-\frac{1}{2}T}^{t_0+\frac{1}{2}T} u\,dt,\quad \overline{v}=\frac{1}{T}\int_{t_0-\frac{1}{2}T}^{t_0+\frac{1}{2}T} v\,dT,\quad \overline{w}=\frac{1}{T}\int_{t_0-\frac{1}{2}T}^{t_0+\frac{1}{2}T} w\,dt.$$

Thus the average values $\overline{u}$, $\overline{v}$ and $\overline{w}$ are in general functions of the position coordinates x, y and z, the time t and the period of averaging T.

At any time t_0 the instantaneous velocity u will be different from its value $\overline{u}$ averaged over the period T; the components of the velocity fluctuation — called eddy or turbulent velocity — along the coordinate axes are

$$u'=u-\overline{u},\quad v'=v-\overline{v},\quad w'=w-\overline{w}.$$

If conditions are such that the average velocity is stable, i.e., $\overline{u}$, $\overline{v}$ and $\overline{w}$ do not change with time, then for such a flow we have

$$\overline{u'}=\frac{1}{T}\int_{t_0-\frac{1}{2}T}^{t_0+\frac{1}{2}T}(u-\overline{u})\,dt=0.$$

Analogously $\overline{v'}=\overline{w'}=0$, i.e., in a steady flow the average eddy velocity is zero.

A strongly developed turbulent flow can be treated as a motion in which a very large number of individual eddies are irregularly distributed. In the early stages of development of the theory of atmospheric turbulence this fact made it possible to regard elementary masses of air as "particles" which behave like molecules and transmit their characteristic properties (heat, motion, specific humidity and the various admixtures contained inside them) from layer to layer.

It proved possible to construct a kinetic picture of turbulence using the fundamental ideas of the kinetic theory of gases. Irrespective of the degree of verisimilitude of this analogy, there can be no doubt that turbulence leads to mixing on a large scale and that as a result the values of the viscosity coefficient obtained from observations of motions in the atmosphere are 10^4–10^5 times larger than the values determined in laboratory experiments with laminar motions. This is to be attributed to the fact that in laminar motions in the laboratory viscosity is induced by the irregular thermal motion of individual molecules (molecular viscosity), whereas in the atmosphere viscosity results from the incomparably greater interaction between [the above-mentioned] particles of the medium [eddies] in turbulent motion, or eddy viscosity (which averages μ_e = = 10–200 g/cm·sec).

Similarly, calorimetric eddy conductivity has a value of the order of $\lambda_e \approx 8-10$ cal/cm · deg · sec as against $\lambda_{mol} = 5.6 \cdot 10^{-5}$, and thermometric eddy conductivity a value of the order of $a_e \approx 30-45 \cdot 10^3 \cdot cm^2 \cdot sec$.

The most important consequence of atmospheric turbulence is the verical and horizontal exchange of air, responsible for mixing the individual gases present in the air and for the constancy of composition. At the same time mixing in the atmosphere enables the transport of any physical quantity. This quantity can be either a certain impurity, e.g., dust, moisture, etc., or the value of some parameter such as momentum or heat content.

§ 2. Coefficient of turbulence

Owing to its importance the study of turbulence is the object of a great deal of attention in meteorology. Apart from the researches of scientists outside the USSR (Reynolds, Prandtl, Taylor, Sutton et al.), very considerable contributions to the development of the theory of atmospheric turbulence have been made by Soviet investigators (A. A. Fridman, L. V. Keller, A. N. Kolmogorov, A. M. Obukhov, M. I. Yudin, D. L. Laikhtman, et al.). Complicated theories of turbulence are given in detail in monographs on dynamic meteorology. Here we shall confine ourselves to deriving certain fundamental relations. In order to obtain the desired relations characteristic of turbulence, let us consider the vertical transport of some physical quantity q in the atmosphere. By the flux Q of the given quantity we will understand the amount of this quantity travelling per unit time (1 sec) across a unit area (1 cm^2) in a direction normal to the area.

Let us take a horizontal surface of s cm^2 at some level z. A great number of air particles will cross this surface per unit time. We will assume that these particles differ from each other only in the value of the quantity q. We denote the value of the quantity q for the j-th particle by q_j and the transverse cross section of the latter by s_j. Then the total amount of the given quantity transferred per unit time across the surface is obviously given by

$$\sum_{j=1}^{N} q_j \rho_j w_j s_j, \tag{1}$$

and the mean flux by

$$\frac{1}{s} \sum_{j=1}^{N} q_j \rho_j w_j s_j. \tag{2}$$

Here N is the total number of moving particles, ρ_j the density and w_j the vertical component of velocity for the j-th particle.

On the basis of what was said earlier about turbulent motions, the total flux through an isolated surface can be treated as the sum of the mean total transport of all particles and the turbulent flux due to the presence of fluctuations (deviations from the mean) of the velocity, i.e., $Q = \overline{Q} + P_{q,e}$, where $P_{q,e}$ is the turbulent flux of the quantity q.

The mean total transport is

$$\overline{Q} = q\rho\overline{w}_z,$$

where ρ is the mean density and $\overline{w}_z$ the mean velocity along the vertical.

Therefore for the vertical turbulent flux of the quantity q one can write

$$P_{q,e} = Q - q\rho\overline{w}_z \tag{3}$$

or, if one introduces the momentum densities $\rho\overline{w}_z = \mu_z$ and $\rho_i w_i = \mu_i$,

$$P_{q,e} = \frac{1}{s}\sum_{j=1}^{N} q_j \mu_j s_j - q\mu_z. \tag{4}$$

This expression could be used for the direct determination of the flux $P_{q,e}$ if the simultaneous values of q_j and μ_j were known for all particles. In (4) the flux is expressed in terms of the sum of a very large number of components of essentially random values; furthermore, the regularities of turbulent flow are not expressed at all explicitly. The discovery of these regularities is indeed the goal of the theory. However, the grave difficulties which lie in the way have not been conclusively solved to this day.

We will transform the expression obtained for the turbulent flux of the quantity q so as to express the physical factors governing the transport explicitly. To do this we assume that the mean total transport is zero, i.e., that $q\rho\overline{w}_z = q\mu_z = 0$ (this is equivalent to assuming that the surface s under consideration moves in the direction of transport at the average speed of the latter). The expression (4) for $P_{q,e}$ then becomes

$$P_{q,e} = \frac{1}{s}\sum_{j=1}^{N} q_j \mu_j s_j. \tag{4'}$$

Let us class all the air particles crossing the surface s at a certain instant t into two groups: 1) those moving upward, with a velocity taken to be positive, and 2) those moving downward, with a negative vertical velocity.

The vertical components of the total momentum of the particles in each group will be equal in value but opposite in sign. In this case, at the given level z and time t we obtain for any particle

$$q_j(z, t) = \overline{q}(z, t) + q_j'. \tag{5}$$

For each such particle one can determine a certain previous instant in time $t_j < t$ at which the particle velocity w_j vanished for the last time (before crossing the level s). At this instant the j-th particle occurred at the level z_j, which we will term initial. For this level

$$q_{0,j}(z_j, t_j) = \overline{q}(z_j, t_j) + q_j''. \tag{6}$$

In expressions (5) and (6) the $\overline{q}$ denote the averages of q for all particles at the levels z and z_j at the corresponding instants t and t_j, while q_j' and q_j'' designate the deviations from the averages.

Let us now expand the quantity $\overline{q}(z_j, t_j)$ in Taylor series, assuming that the differences $(z_j - z)$ and $(t_j - t)$ are so small that terms of the second and higher orders can be disregarded, i.e., we write

$$\overline{q}(z_j, t_j) = q(z, t) + (z_j - z)\frac{\partial\overline{q}(z, t)}{\partial z} + (t_j - t)\frac{\partial\overline{q}(z, t)}{\partial t} \tag{7}$$

or

$$\overline{q}(z_j, t_j) - q(z, t) = (z_j - z)\frac{\partial\overline{q}(z, t)}{\partial z} + (t_j - t)\frac{\partial\overline{q}(z, t)}{\partial t}. \tag{7'}$$

We denote by δq_j the change in the quantity q_j for the j-th particle as it moves from the level z_j to the level z. Then

$$\delta q_j = q_j(z, t) - q_{0,\,j}(z_j, t_j). \tag{8}$$

In view of (6) and (7) we obtain

$$q_j(z, t) = \delta q_j + \bar{q}(z_j, t_j) + q_j'' =$$
$$= \delta q_j + \bar{q}(z, t) + (z_j - z)\frac{\partial \bar{q}(z, t)}{\partial z} + (t_j - t)\frac{\partial \bar{q}(z, t)}{\partial t} + q_j''. \tag{9}$$

After introducing the obtained expression into (4') we have the following expression for the turbulent flux

$$P_{q,\,e} = \bar{q}(z, t)\frac{1}{s}\sum_1^N \mu_j s_j + \frac{\partial \bar{q}(z, t)}{\partial z}\frac{1}{s}\sum_1^N \mu_j s_j (z_j - z) +$$
$$+ \frac{\partial \bar{q}(z, t)}{\partial t}\frac{1}{s}\sum_1^N \mu_j s_j (t_j - t) + \frac{1}{s}\sum_1^N \mu_j s_j \delta q_j + \frac{1}{s}\sum_1^N \mu_j s_j q_j''. \tag{10}$$

Let us now analyze this expression. The first term is rigorously zero on the basis of our assumption that the mean flux is zero. The third term can also be disregarded if one considers that the sums of the positive and negative components in it should approximately compensate each other. Disregarding differences in density and assuming that $\mu_i = \rho_i w_i \approx \rho w$, one can write instead of (10)

$$P_{q,\,e} = -\frac{\rho}{s}\frac{\partial \bar{q}(z, t)}{\partial z}\sum_1^N w_j s_j (z - z_j) + \frac{\rho}{s}\sum_1^N w_j s_j \delta q_j +$$
$$+ \frac{\rho}{s}\sum_1^N w_j s_j q_j''. \tag{11}$$

Here the two last terms reflect the influence of vertical motion on the variation of the value of q in the particles and the inverse influence of the value of q on the particle displacement. It appears that in the transport of the majority of physical quantities which are meteorologically important these terms can be disregarded in the first approximation, i.e., one can consider that the process of variation of the values of q in the particles is not related to the character of the motion of the latter. Then expression (11) for vertical turbulent flux takes on the simple form

$$P_{q,\,e} = -\rho k_z \frac{\partial \bar{q}}{\partial z}, \tag{12}$$

where k_z denotes the quantity

$$k_z = \frac{1}{s}\sum_{j=1}^N w_j (z - z_j) s_j, \tag{13}$$

which is called the coefficient of turbulence [eddy viscosity]. This quantity characterizes the intensity of transport of various physical quantities in the z-direction as a result of turbulence. Since z_i is the level at which w_i is zero, for particles approaching the level z from above $w_j < 0$ and $(z - z_j) < 0$, whereas for particles approaching from below $w_j > 0$ and

$(z-z_j)>0$; therefore for all particles the product $w_i(z-z_i)$ is always greater than zero and, consequently, k_z is a positive quantity.

The relation (12) can be written in the form

$$P_{q,\,e}=-k_z\rho\frac{\partial\bar{q}}{\partial z}=-A_z\frac{\partial\bar{q}}{\partial z}, \tag{14}$$

where

$$A_z=k_z\rho. \tag{15}$$

The quantity A_z is called the coefficient of turbulent exchange or mixing coefficient. The coefficient of turbulent exchange A is numerically equal to the flux of the quantity q for a unit concentration gradient $-\frac{\partial\bar{q}}{\partial z}$.

Since the dimensionality of the flux is $g/cm^2\cdot sec$ and that of the gradient cm^{-1}, the dimensionality of the coefficient of exchange is obviously $g/cm\cdot sec$ and that of the coefficient of turbulence cm^2/sec (or m^2/sec).

The turbulent fluxes of the quantity q in the horizontal direction (along the x and y axes) can be represented analogously:

$$P_{x,\,e}=-\rho k_x\frac{\partial\bar{q}}{\partial x}\quad\text{and}\quad P_{y,\,e}=-\rho k_y\frac{\partial\bar{q}}{\partial y}, \tag{16}$$

where k_x and k_y are the coefficients of turbulence in the horizontal direction.

It should be stressed that the coefficients of turbulence do not reflect any peculair property of the quantity q and simply represent a kinematic characterization of the properties of turbulent transport; therefore the turbulent fluxes of widely different physical quantities are expressed in terms of the same coefficients k.

At the same time, since they are kinematic characteristics, the coefficients of turbulence are heavily dependent on the rate of flow, the proximity of solid bodies, temperature gradients, wind speed and many other factors.

Expression (13) for the coefficient of turbulence can be represented in a different form. Indeed, since at the level z_j the velocity w_j is zero for every particle, one can assume with a sufficient degree of accuracy that w_j is proportional to $(z-z_j)$. The proportionality factor should obviously have the dimensionality of the velocity gradient. L. Prandtl suggested writing it in the form $\alpha_j\left|\frac{\partial\bar{u}}{\partial z}\right|$, where α_j is a certain dimensionless factor. Therefore following Prandtl one can write

$$w_j=\alpha_j(z-z_j)\left|\frac{\partial\bar{u}}{\partial z}\right|,$$

and then we obtain instead of (13)

$$k_z=\frac{1}{s}\sum\alpha_j(z-z_j)^2 s_j\left|\frac{\partial\bar{u}}{\partial z}\right|. \tag{17}$$

Setting

$$\frac{1}{s}\sum\alpha_j(z-z_j)^2 s_j=l^2, \tag{18}$$

we finally find

$$k_z=l^2\left|\frac{\partial\bar{u}}{\partial z}\right|. \tag{19}$$

The quantity l introduced by the relation (17) has the dimensionality of length and is called the mixing length.

The introduction of the mixing length sometimes proves to be very convenient, especially when the mixing length can be determined from simple physical relations. From expression (17) it is evident that if the factor α_j could be set equal to unity the mixing length would be the root-mean-square distance between the initial level at which the eddy appears and the final level at which it breaks down and merges with the surrounding medium.

More detailed investigations show that the coefficient of turbulence k_z increases linearly with altitude in the lower layer (to a height of several meters). At greater altitudes its vertical increase can be expressed by an exponential law, but in practice, starting from a height of a few tens of meters, the increase becomes so slight that k_z can be regarded as constant. As a result in many investigations the coefficient of turbulence is represented by a function of height with a discontinuity at a certain height (Figure 6).

That layer of the atmosphere in which the coefficient of turbulence increases with height is called the surface layer. It extends up to a height of several tens of meters, most frequently to 50–100 m*.

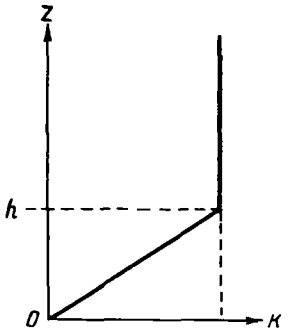


FIGURE 6. Variation of the coefficient of turbulence with height

The expression obtained for turbulent flux (14) is completely analogous in form to the expression for the molecular flux Q_{mol}

$$Q_{mol} = -\rho D \frac{\partial q}{\partial z}, \qquad (20)$$

where D is the coefficient of molecular diffusion or exchange, which has the same dimensionality as k; the molecular flux, moreover, is also proportional to the gradient of the concentration.

The coefficient of molecular diffusion, which is strongly dependent on the temperature, is 0.2 cm^2/sec at 0°, whereas the coefficient of turbulence is tens and hundreds of thousands times larger ($(1.5-40)\cdot 10^3$ cm^2/sec). Consequently, the magnitude of the turbulent flux is also tens and hundreds of thousands of times greater than the magnitude of the molecular flux.

It should be stressed here that this analogy between expressions (14) and (19) is purely external and in no way reflects the essential difference in their physical meaning.

A great many methods have now been elaborated for the determination of the coefficient of turbulence. The exchange coefficient in the surface layer is determined on the basis of so-called gradient observations with the use of formulas relating the coefficient of turbulence to the variation of various physical quantities (wind velocity, temperature and others) with height. These methods are described in detail in the appropriate instructions; we will consider a few later on. Such observations reveal, as we noted above, that the coefficient of turbulence in the atmosphere varies sharply in time and space; it depends on the vertical gradient of wind velocity and temperature, on the thermal inhomogeneity of the underlying surface and on many other factors.

* [Often called the "lower atmosphere" in Western usage. However, in Russian usage the latter term has a different meaning. See p. 50.]

For a very broad range of values of k_z, varying roughly from values of the order of the coefficient of molecular diffusion up to 10^4–10^5 cm^2/sec, the mean values of k_1 at a height of 1 m, under daytime conditions in the warm part of the year and on land, are of the order of 1500–2000 cm^2/sec. The values of k_1 depend strongly on the distribution of temperature with height and wind velocity, which determines the daily and annual march of k_1. In the daily march, which is particularly clear-cut in fine (cloudless) weather, the maximal values in the warm part of the year are observed in the afternoon hours and the minimal values in the night and early morning. This is

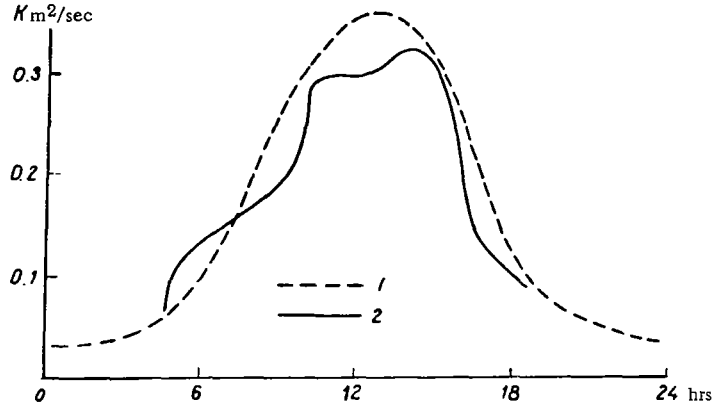


FIGURE 7. Daily march of the coefficient of turbulence

1-calculated; 2—after M. P. Timofeev.

illustrated in Figure 7, which shows the variation of k_1 over the day at lat. 60°N. from calculated (1) and experimental (2) data. In the annual march for middle latitudes the exchange coefficient has maximal values in summer and minimal values in winter. Over extensive bodies of water, where vertical gradients of air temperature are comparatively limited, the exchange coefficient in the air layer immediately above the water depends on the wind velocity and partly on the degree of agitation.

Regarding the horizontal mixing of air, it is now known that its intensity is comparatively independent of the height above the ground, and is proportional to the velocity of the air flow and to the size of the eddies which appear. If one treats individual cyclones and anticyclones as turbulent formations one arrives at values of the order of 10^6 m^2/sec for the horizontal exchange coefficient.

§ 3. Equation of turbulent mixing

In order to determine the variation in time of a certain quantity q owing to turbulent exchange, we will consider an element of volume of height dz

and cross section 1 cm^2 lying at the height z. The flux of this quantity through the lower surface of the volume element is

$$Q_z = -A_z \frac{\partial q}{\partial z}, \tag{21}$$

and through its upper surface (at the height $z + dz$)

$$Q_{z+dz} = -A_{z+dz}\left(\frac{\partial q}{\partial z}\right)_{z+dz} = -A_z \frac{\partial q}{\partial z} - \frac{\partial}{\partial z}\left(A_z \frac{\partial q}{\partial z}\right)dz =$$
$$= -\left[A_z \frac{\partial q}{\partial z} + \frac{\partial A_z}{\partial z}dz \frac{\partial q}{\partial z} + A_z \frac{\partial^2 q}{\partial z^2}dz\right]. \tag{22}$$

Subtracting (21) from (22), we obtain the following expression for the change in the quantity in the given volume per unit time

$$dQ = Q_{z+dz} - Q_z = -\left[\frac{\partial A_z}{\partial z}\frac{\partial q}{\partial z} + A_z \frac{\partial^2 q}{\partial z^2}\right]dz. \tag{23}$$

Since the mass of the volume in question is ρdz, the increase in the quantity q per unit time and unit mass is given by

$$\frac{dQ}{\rho dz} = \frac{\partial q}{\partial t} = -\frac{1}{\rho}\left[\frac{\partial A_z}{\partial z}\frac{\partial q}{\partial z} + A_z \frac{\partial^2 q}{\partial z^2}\right] = -\frac{\partial}{\partial z}\left[\frac{A_z}{\rho}\frac{\partial q}{\partial z}\right] \tag{24}$$

and, if A_z is independent of z,

$$\frac{\partial q}{\partial t} = -\frac{A}{\rho}\frac{\partial^2 q}{\partial z^2}. \tag{25}$$

The above is the equation of turbulent mixing which is widely applied in meteorology.

In deriving the above relations, as noted earlier, it was understood (1) that the quantity in question was conservative, i.e., its value in each elementary particle of air was conserved during its motion and in the displacement of two air particles the values of q were simply added, and (2) that its admixture to the air did not affect the character of the motion of the air particles — the quantity was "passive"; for this reason the last two terms in equation (11) are dropped.

In cases where these two conditions are not satisfied the problem, as we shall see later on, becomes far more complicated.

Chapter 4

STRUCTURE OF THE ATMOSPHERE

§ 1. Methods of investigation of the upper atmospheric layers

The need for atmospheric research in the form of observations at different heights above the ground was already realized in the eighteenth century. In 1754 at a meeting of the Academy of Sciences M. V. Lomonosov demonstrated an "aerodromic" machine invented by him, designed to "lift thermometers and other small meteorological instruments".

With the birth of aerostatics meteorological observations were carried out in practically all balloon ascents. However, methods of free-atmosphere research developed slowly; the basic aerological methods which have made it possible to observe the free atmosphere regularly by comparatively simple means were evolved only around the turn of the present century. Basically these methods began to develop subsequent to the invention of the radiosonde by P. A. Molchanov in 1930.

As aviation developed the airplane was also employed in probing the atmosphere. Today specially-equipped flying meteorological laboratories are installed on aircraft. However, the height to which information concerning the atmosphere can be obtained by such means is comparatively limited. For the radiosonde method the limit is 20–25 km, and only in a few isolated instances do radiosondes attain a height of 30–35 km. The range of sounding by aircraft is determined by the aircraft ceiling and is likewise limited. The application of the ideas of K. E. Tsiolkovskii, the scientist responsible for the development of rocket technology, made it possible to use special geophysical and meteorological rockets in upper-atmosphere studies (Figure 8). One rocket, launched in the USSR on 21 February 1958 and weighing 1520 kg, attained a record height for single-stage rockets in this class — 473 km above the ground. On 27 August of the same year a rocket weighing 1690 kg ascended to an altitude of 450 km.

One feature of geophysical rockets is that at a predetermined height they eject containers carrying all the required scientific apparatus; these are then lowered to the surface by parachute. In the meteorological rockets of the Central Aerological Observatory the measuring equipment is placed in the nose of the rocket, which is also lowered by parachute.

The launching of cosmic rockets and artificial satellites has made it possible to penetrate to yet higher altitudes and to obtain valuable geophysical information concerning not only the earth's atmosphere but also those regions of cosmic space closest to the earth. Completely reliable data can be obtained with the aid of meteorological rockets to a height of

about 80–100 km. For greater altitudes data become less accurate and the number of ascents is smaller.

It should be mentioned that the collection of data by rockets and satellites, in addition to representing a complex technical problem, also requires the development of special equipment. Owing to high rocket speeds and the high degree of rarefaction, far from all physical characteristics of the state of the atmosphere are directly measurable (especially in the upper layers). Even such important characteristics as air temperature and density are usually calculated from measurements of certain parameters.

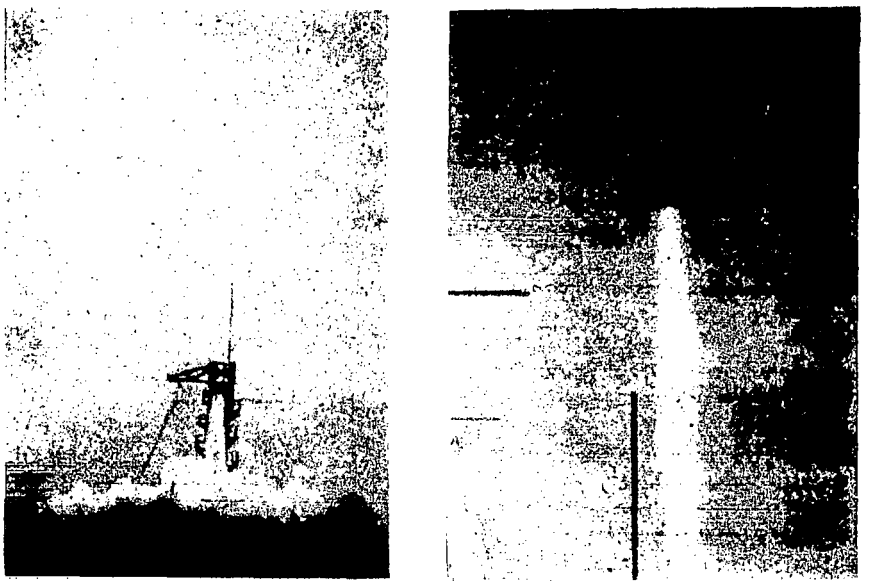


FIGURE 8. The launching of a meteorological rocket

In addition to the experimental methods for studying the upper layers of the atmosphere so-called indirect physical methods are also used; these may be termed passive or active according to the method used.

Active methods are those which involve observation of the influence exerted by the atmosphere on light rays directed upwards (projector method), sound waves formed in large explosions (acoustic method), and, finally, the propagation of radio waves (radio physical methods). By investigating refraction, reflection, scattering, polarization and other processes observed in the propagation of these waves and applying theoretical considerations, one can arrive at important data concerning the state of the atmosphere up to the altitudes penetrated by these waves.

Passive methods of upper-atmosphere research are based on the study of phenomena which arise naturally in various layers of the atmosphere and the character of which depends on the state of these layers. These phenomena include, for instance, the so-called mother-of-pearl clouds observed at heights of 22–27 km, as well as the noctilucent clouds which occasionally

appear at twilight and vanish after twilight is over. They occur at altitudes between 80 and 85 km (82 km on the average) and shine with a silvery, slightly bluish, fairly bright light. Although their origin is still not entirely clear, there is reason to believe that they are composed of water droplets or ice crystals as are ordinary clouds. The speed of air currents at these altitudes can be estimated from their rate of motion.

Important data on the density and temperature of the upper layers can be obtained by observing meteors (falling stars), cosmic bodies varying in size from minute (mass of several milligrams) to very large in rare cases. The bright light which traces the meteor's fall from the moment of ignition to its extinction creates the illusion of a swiftly falling and fading star. By observing the altitude of ignition and extinction, and also the displacement of so-called meteor trails (luminous formations at night and dust traces in daytime), it is possible to obtain valuable data concerning the density of the atmosphere and air motions at these altitudes.

Still more extensive data can be obtained by optical observation of such phenomena as twilight, nightglow and auroras.

Spectroscopic, spectrophotometric and polarization observations in conjunction with theoretical data make it possible to form some idea of the composition of the atmosphere at high altitudes; they also provide a picture of dissociation, excitation, ionization and many other processes taking place in the rarefied gaseous medium at these heights.

Most of the methods mentioned here will be considered in later sections devoted to the description of these phenomena.

One may safely assume that present information concerning the atmosphere to a height of the order of 100 km is fully reliable and definite, though not always complete. For higher layers our information is at present limited and less accurate. As regards the uppermost atmospheric layers (500 km), the data are still hypothetical in character and only a more or less probable overall scheme of their structure can be given. However, investigations with the help of rockets and satellites provide new data daily; frequently these are not entirely in accord with earlier ideas, which are therefore continually being modified and corrected.

§ 2. Stratification of the atmosphere

The atmosphere is not homogeneous throughout. Its state and properties vary significantly with altitude as well as from point to point. The variation of its properties is particularly marked along the vertical. The atmosphere may be divided vertically into a number of layers differing in composition, temperature regime, electrical characteristics and other physical properties. The contrast between these layers, the largest of which are called spheres, is particularly conspicuous in the vertical distribution of temperature. As a result since the beginning of the century it has been the practice in meteorology to divide the atmosphere into two parts: the troposphere (lower layer) and the stratosphere, by which was originally meant all atmospheric layers lying above the troposphere.

The troposphere is the lowermost layer of the atmosphere; it is a thin one compared to the height of the entire atmosphere but contains the bulk of its mass — about three-quarters of it. The upper limit of the

troposphere lies at a height ranging from 7 to 18 km depending on latitude, time of year and properties of the earth's surface; its mean altitude in middle latitudes is taken to be roughly 10–11 km. The most characteristic feature of the troposphere is a falling-off of temperature with altitude, by about 6° per 1 km (on an average). The troposphere is a region of particularly strong vertical mixing and heat exchange with the earth's surface, and the properties of the latter exercise a considerable influence on all processes taking place inside it. It contains practically all the water vapor; all the clouds usually observed are formed within it; lastly, it is the site of the principal weather processes. In the present book our attention will therefore center around phenomena occurring in the troposphere.

The layers above the troposphere which were formerly referred to jointly as the "stratosphere" are now subdivided into a number of spheres the names and mean altitudes of which are indicated in Table 8.

TABLE 8

Major layers (spheres) of the atmosphere

Name	Mean height (km)	Transitional layer	Mean height (km)
Troposphere	0–10	Tropopause	10–11
Stratosphere	11–50	Stratopause	50–55
Mesosphere	55–80	Mesopause	80–85
Thermosphere	85–500	Thermopause	-
Exosphere	above 500		

Note: Heights refer to middle latitudes.

The transition from each of these spheres to the next is always effected through an intermediate layer, more or less thick and also displaying a number of peculiarities. These transitional layers are known by special names formed by replacing the suffix "sphere" in the name of the layer below by the suffix "pause". Thus the intermediate layer which is best known, namely from the troposphere to the stratosphere, is called the tropopause; similarly one can speak of the stratopause, the mesopause and the thermopause.

The stratosphere is distinguished by the fact that its temperature remains nearly constant, or increases, with altitude. Its temperature minimum (and also the minimum temperature of the tropopause) is observed in the equatorial region, where it is −70 to −80°. In middle latitudes it is approximately −55 to −60°.

Starting from a height of about 35 km the temperature rises considerably and at the stratopause (50–55 km) reaches about 0°. This temperature rise is due to the absorption of solar radiation in the ozone layer which is found at these altitudes. Above 55 km lies the mesosphere, characterized by a falling-off of temperature which continues until the mesopause (80–85 km), where the temperature drops to −70°. This is followed by the transition from the mesosphere to the next layer, called the thermosphere. This is the thickest layer of all, characterized by a continuous rise in temperature with altitude up to its upper boundary, which occurs at a height of several hundred kilometers. Above the thermosphere

lie the external layers of the atmosphere or exosphere; the temperature there does not change with altitude, or possibly increases somewhat, and is very high (over 1000–1500°K). This outermost sphere is distinguished by the fact that it is the region where gases escape from the earth's atmosphere into interplanetary space. It extends to a height of 2000–3000 km, where the atmosphere comes into contact with interplanetary gases.

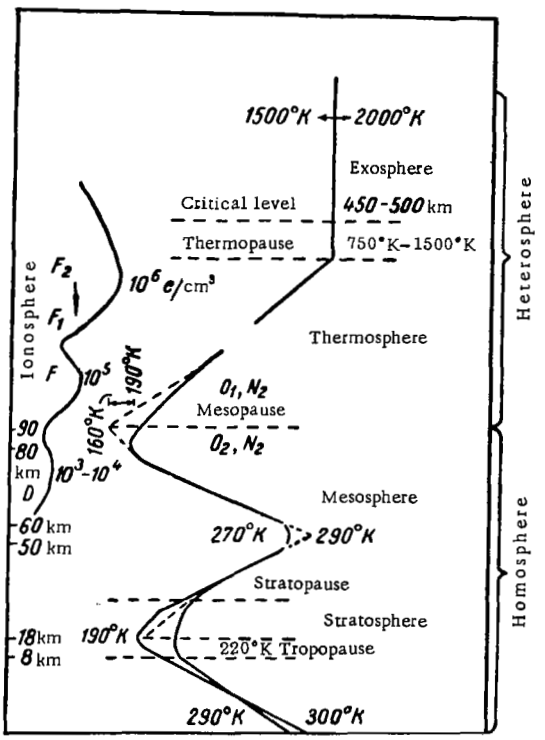


FIGURE 9. Vertical structure of the atmosphere, after Kuiper

The thermosphere and especially the exosphere have been studied far less than the layers situated below. Their distinguishing feature is an exceedingly low air density coupled with the fact that, under the strong influence of cosmic radiation and ultraviolet rays from the sun, various processes — excitation, dissociation, ionization and the converse processes of emission, recombination, etc.— take place intensively. This explains the development in these layers of such phenomena as high-conductivity layers, polar auroras, nightglow and so forth.

In view of these differences between the troposphere, on the one hand, and the thermosphere and exosphere, on the other hand, one often speaks of the lower atmosphere, meaning the troposphere and lower stratosphere, and the upper atmosphere, meaning layers situated above 30–40 km.

In addition to the major layers differentiated by the criterion of temperature, the term ionosphere is used to indicate a layer characterized by high electrical conductivity of the air and the presence of a large number of ions and electrons (see Ch. 30). The lower edge of the ionosphere occurs at a height of about 60–80 km, and it rises up to a height of several hundred kilometers.

We note further that in the troposphere, stratosphere and mesosphere (0–80 km) the composition of the atmosphere with reference to the principal gases (and therefore also the mean molecular weight of air) remains roughly the same despite the presence of small amounts of certain supplementary gases (notably ozone). It has therefore been suggested that this region of the atmosphere be called homosphere, as distinct from the overlying part or heterosphere, where the molecular weight of air varies with altitude as a result of the processes of dissociation of oxygen and nitrogen.

The subdivision of the atmosphere into layers is illustrated schematically in Figure 9.

§ 3. Vertical extent of the atmosphere. The escape (dissipation) of gases from the atmosphere

When we speak about the atmosphere the question of its vertical extent or the height of its upper boundary arises naturally. Since the density of air decreases continuously with altitude and the atmosphere grades into interplanetary space, it has no naturally well-defined upper bound. The latter can only be established arbitrarily from certain physical data and theoretical considerations. For example, the upper boundary can be taken to be the height to which the presence of the atmosphere is confirmed by the observation of certain physical phenomena. This so-called meteorological height can be taken as maximally equal to 1000–1200 km, a level at which polar auroras are still observed occasionally.

The upper limit of the atmosphere can also be established from other considerations. The molecules and atoms of the gases which compose the atmosphere occur in a state of random (thermal) motion, the average velocity of which increases with temperature. As altitude increases and the density of the atmosphere decreases the number of molecules becomes increasingly small; this reduces the probability of collision between particles and lengthens the mean free path of the atoms. Under such conditions at a certain altitude it might happen that a particle to which an ascending motion had been imparted in collision at some moment would encounter no other particles on its way, its motion being subject solely to the influence of gravity.

Since the kinetic energy of a molecule $W_{kin} = \frac{mv^2}{2}$, where m is the mass of the molecule and v its velocity, and the potential energy in the earth's gravitational field $W_{pot} = \int_z^\infty mgdz$, where g is the acceleration of gravity, which is equal to $\frac{\gamma M}{r^2}$ (γ being the constant of gravitation, M the mass of

the earth and r the distance of the molecule from the center of the earth), it is obvious that if W_{kin} is greater than W_{pot} the molecule will be able to fly into interplanetary space and escape from the earth's atmosphere. This phenomenon is called the escape of atmospheric gases, or dissipation of the atmosphere. In order for escape to be possible it is necessary that the thermal velocity of the molecule reach a certain critical value v_{cr}, which can be found from the inequality $W_{kin} \geqslant W_{pot}$ and can be expressed by the relation

$$v_{cr} \geqslant \sqrt{\frac{2\gamma M}{r}}.$$

The numerical value of v_{cr} is approximately 11 km/sec (at a height of 800 km above the surface $v_{cr} = 11.85$ km/sec).

The height at which the escape of gases becomes possible is known as the critical level.

The complex question of the escape of gases from the earth's atmosphere (and also from the atmospheres of other planets) has been studied by many investigators who have sought to determine the intensity (rate) of escape and calculate the height at which it begins.

We recall that the root mean square velocity of molecules is given by the relation $v = \sqrt{\frac{3kT}{m}}$, where k is Boltzmann's constant and m the mass of the molecule. Consequently, v depends on the temperature T of the gas and on its molecular weight. Let us also recall for any gas owing to the Maxwell velocity distribution of the molecules some of its molecules will always have velocities higher than the critical one. Taking this into account we arrive at the conclusion that, for any gas, the higher its temperature and the lighter it is, the greater the number of molecules escaping from the atmosphere. In the earth's atmosphere such gases are first and foremost hydrogen and helium.

The process of escape of gases takes place very slowly, especially for gases with a large molecular weight. Calculations show that only those gases for which $v_{av} > v_{cr}$ could have escaped entirely from the atmosphere in the period elapsed since the earth came into being (about $3 \cdot 10^9$ years; this is in the absence, of course, of constant sources of replenishment). The rate of escape of the lightest gases may be deduced from the tentative data for the time of complete dissipation of hydrogen and helium given in Table 9.

TABLE 9

Complete dissipation time τ of hydrogen and helium

	Helium (v_{cr} = 1300 m/sec at 0°)			Hydrogen (v_{cr} = 1836 m/sec at 0°)		
$T°$	273	500	1000	273	500	1000
τ(years)..	$7 \cdot 10^{40}$	$2.6 \cdot 10^{19}$	$1.4 \cdot 10^{6}$	$1.8 \cdot 10^{7}$	$5.3 \cdot 10^{6}$	4.0

It is obvious from the table that at the high temperatures which doubtless obtain in the upper atmospheric layers hydrogen should escape very rapidly. Helium escapes at a slower rate but the time for its complete dissipation at a temperature of about 1000° is less than the elapsed life of the earth.

This is why helium does not accumulate in the earth's atmosphere despite the fact that it is being constantly released by the crust. It is estimated that the dissipation of the helium liberated by the crust would require a temperature of 1500–2000°K as well as a reduction of the density to a value of roughly $3 \cdot 10^7$ cm^{-3}.

Many authors have found by computation that the height at which these conditions are fulfilled lies between 500 and 1000 km. From this it is concluded that the height of the critical level occurs somewhere between these values. Mitra, for example, considers 800 km to be the most probable height for the lower base of the exosphere.

We might mention another point. Molecules and atoms with velocities less than the critical one may also avoid collision while travelling in the exosphere, and yet fail to leave the atmosphere; instead, describing an elliptical trajectory, they reach a certain great altitude, determined by their velocity and the acceleration of gravity at the point of impact, and then fall back. The time spent by the atom in this ascent and return to the critical level is insignificant, i.e., of the order of a few minutes. Thus we are led to conclude that there is a region starting at a certain level in the atmosphere in which collisions between atoms can be disregarded. Atoms which, in the course of their ascent, enter this region from below, fall downward once again; only those atoms whose velocity is greater than the critical velocity will be able to escape from the earth's atmosphere. The region in which this takes place, as we noted earlier, is called the exosphere (i. e., outer sphere).

According to the most reliable estimates one may assume that the exosphere occurs on the average somewhere between 500 and 1000 km (near 800 km; it begins at the height of the critical level) and extends to 2000–2500 km, at which height the density (number of particles per cubic centimeter) tentatively reaches $1/cm^3$; this corresponds to the density of interplanetary gas.

§ 4. Hypotheses concerning the origin and evolution of the atmosphere

Closely related to the question of dissipation is the question of the origin of the atmosphere and the possible changes in its composition over geologic time.

Astrophysicists have established the probable composition of the atmospheres of other planets in the solar system. It has been shown that the composition of the earth's atmosphere differs essentially from that of other planets (especially large ones), which have atmospheres consisting principally of hydrogen, helium and methane.

The question of the formation and evolution of planetary atmospheres was first stated as a problem of considerable cosmogonic significance by Academician V. G. Fesenkov. The final solution to this problem is, however, still remote. The same can be said of the origin of the earth's atmosphere and the evolution of its composition in the earth's elapsed life. We shall nonetheless dwell briefly on this question.

Obviously, the formation of the primary atmosphere was conditioned by the process of formation of our planet. According to the hypothesis which holds that the planets of the solar system evolved from molten matter

ejected by the sun, the original atmosphere must have consisted of the same light, incandescent gases as the sun's atmosphere. Owing to the high temperature and small mass of the earth this atmosphere was swiftly scattered into interplanetary space. A secondary atmosphere was created as a result of the cooling of the earth and of the various chemical reactions taking place in it.

Today the accepted hypothesis is that of the "cold" origin of the planets put forward by Academician O. Yu. Shmidt. According to this hypothesis planetesimal material generated the principal gases composing the atmosphere. The subsequent process of warming of the earth also led to the dissipation of light gases from the primary atmosphere.

The second question, that of the possible alterations in the composition of the atmosphere in the course of geologic time, can be approached on the basis of geologic data concerning the composition of rocks, the process of their formation, and their gaseous content, as well as by calling upon data relating to the distribution of chemical elements and their content on earth and in cosmic space.

At present it seems one may state that even in the distant geologic past climatic and atmospheric conditions were not widely different from today's conditions, and that the atmosphere always contained a fairly large amount of oxygen and comparatively small amount of carbon dioxide. This does not exclude, of course, the cyclic changes over extensive periods of time (hundreds of thousands and millions of years) which have indisputably taken place on earth and during which the atmosphere has periodically changed composition to some extent. But these changes were probably not very great.

It may be assumed that nearly all the gases present in the atmosphere enter it by way of the earth's crust, chiefly during volcanic eruptions. From the atmosphere they return by complex ways to the earth's surface, as in the case of carbon dioxide, which is used up in plant respiration and in the formation of sedimentary rocks. The marked prevalence of nitrogen in the atmosphere can be explained by the fact that after its generation in the initial stage of formation of the atmosphere it was preserved in much greater amounts than the other gases owing to its relatively high inertness.

The presence of large amounts of free oxygen in the atmosphere is very difficult to explain, since free oxygen is not generated in volcanic eruptions and the amount contributed by the interior of the earth cannot be at all significant. This has given rise to the hypothesis of a biogenic origin of oxygen according to which significant quantities are generated as a result of the photosynthetic activity of plants; this view has received a fair degree of confirmation.

§ 5. Distribution of temperature, pressure and density in the upper atmospheric layers

Earlier we gave some idea of the general character of the distribution of temperature with altitude. We will now consider in greater detail the question of the fundamental parameters characterizing the state of the upper layers of the atmosphere. These parameters are temperature, pressure and density. We first note the following. Usually when speaking of the

temperature of a gas we mean its gas-kinetic temperature T, related to the speed of the molecules $\bar{v}$ by $\bar{v}=\sqrt{\frac{8R^*T}{\pi\mu}}$, where R^* is the universal gas constant and μ the molecular weight. This relation is valid under the conditions of a Maxwell velocity distribution; for a sufficiently dense gas the number of collisions between its molecules is large and the mean kinetic energies of all molecules are equal. In this case all the basic gas-kinetic relations can be used to describe the state of the gas.

On the other hand, if the gas is in a state of high rarefaction collisions between particles become rare and the mean free path of the molecules is large. It then becomes impossible to achieve thermal equilibruim between the gas and the thermometric body (thermometer bulb) placed inside it—the number of molecules striking the thermometer will be small and the latter will indicate a temperature completely different from the temperature of the gas surrounding it, as defined by the above relation. This explains the difficulty and even impossibility of using direct thermometric methods to measure the temperature of highly rarefied gases. For this reason the temperature of the upper layers of the atmosphere is usually determined by various indirect methods involving the measurement of parameters related to the gas-kinetic temperature. Thus in rocket research the temperature and density of air are usually computed on the basis of relations derived from the gas-dynamical relation (modified Bernoulli equation) $p_0=3\rho v^2+\frac{1}{2}Bp$, which relates the rocket velocity v to the atmospheric pressure p and the so-called drag pressure p_0, from the equation of state for an ideal gas $p=\frac{R^*}{\mu}\rho T$, and from the fundamental equation of hydrostatics $dp=-\rho g dh$.

A method of direct determination of the temperature by means of resistance thermometers was evolved for investigations with the aid of geophysical and meteorological rockets in the USSR. This method gives an accuracy of ± 20° up to an altitude of 75–80 km. At greater heights, however, even this method becomes unsuitable and the temperature is usually calculated from the measured pressure. But great difficulties arise in such calculations owing to a multiplicity of causes, among them the variation of the molecular weight of air with altitude (above 100 km) due to the dissociation of oxygen and nitrogen.

TABLE 10

Temperature as a function of μ, from rocket data

Altitude (km)	T°K for μ = 28.97	μ	T°K
80	205	28.97	205
100	240	26.22	217
120	330	23.95	273
150	503	20.06	348
200	792	15.79	432

Table 10 provides an idea of the magnitude of the divergences in the values of the temperature as calculated for a constant molecular weight (μ = 28.97) and for values of the molecular weight varying with altitude.

It is evident from the table that the temperature values obtained in the presence of dissociation (decreasing μ) were lower.

Another important question is that of the height to which the Maxwell distribution and the gas-kinetic relations still remain valid. After analyzing this problem Gerson arrived at the conclusion that thermal equilibrium obtains to considerable heights (not less than 400 km) and that one can speak with complete justification of the gas-kinetic temperature of these layers. This equilibrium may indeed be absent at high altitudes.

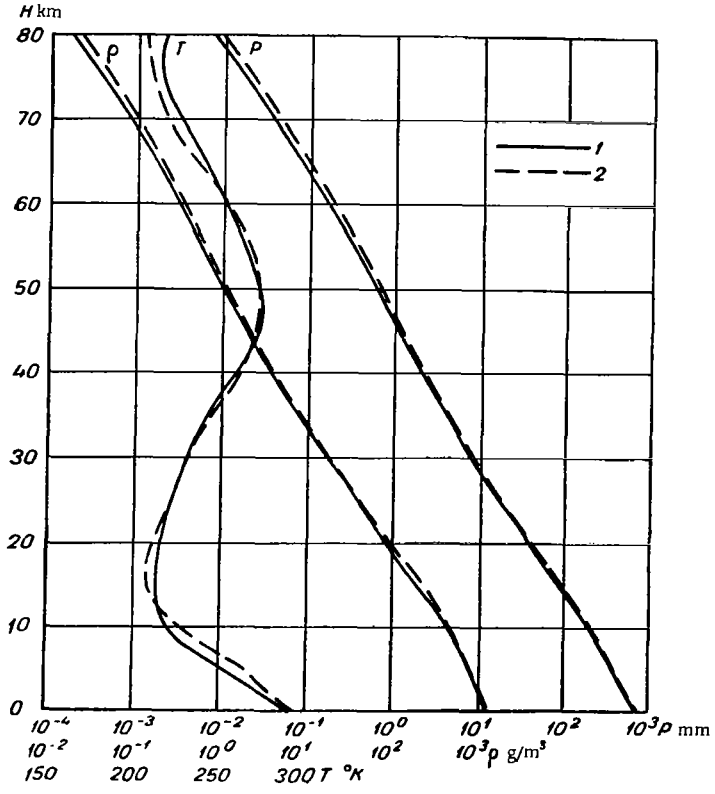


FIGURE 10. Vertical distribution of pressure (p), temperature (T) and density (ρ) according to rocket data

1–TsAO data; 2–US data.

Thus to a height of 100 km results obtained for T by various investigators are fairly close; for greater altitudes results diverge significantly owing to differences in the estimation of initial conditions. For a height of about 300 km, for example, results can differ by several hundreds of degrees (by roughly 500°K and over), while at very great altitudes values are in general purely tentative.

Current ideas on the distribution of temperature, pressure and density in the atmosphere will certainly be subject to continual correction, and the

numerical values cited below should not be regarded as exact or final. Since temperatures at the altitudes in question are determined chiefly by the absorption of ultra-violet radiation from the sun, thermal conditions at different periods may vary. There should therefore be considerable seasonal and daily fluctuations of the temperature. This is supported by observations, which show that even the differences between daytime and nighttime values of T at the same altitude can reach a few hundred degrees. At certain times, in fact, during bursts of ultra-violet emission, temperatures may even reach the very high values (3000–4000°K) cited in certain reports.

In 1947 analysis of the many data obtained by various indirect methods made it possible to unify them and determine the so-called NACA standard atmosphere, recommended by the National Advisory Committee for Aeronautics of the US for various calculations. However, the earliest generalizations of rocket data published in 1952 already revealed themselves to be significantly at variance with the NACA data. The results of rocket research to a height of 80 km in the USSR (TsAO) and USA are shown in the left-hand side of Table 11 and in Figure 10.

TABLE 11

Temperature, pressure and density of the atmosphere

Height (km)	Temperature (°K)	Pressure (mm)	Density (g/m^3)	Height (km)	Temperature (°K)	Pressure (mm)	Density (g/m^3)
0	294	757	1190	90	211	$1.9 \cdot 10^{-3}$	$4.12 \cdot 10^{-3}$
2	278	598	989	100	237	$4.3 \cdot 10^{-4}$	$8.29 \cdot 10^{-4}$
4	261	466	815	110	267	$1.2 \cdot 10^{-4}$	$1.97 \cdot 10^{-4}$
6	247	358	661	120	301	$4.0 \cdot 10^{-5}$	$5.61 \cdot 10^{-5}$
8	233	270	530	130	340	$1.6 \cdot 10^{-5}$	$1.90 \cdot 10^{-5}$
10	220	201	420	140	380	$7.4 \cdot 10^{-6}$	$7.57 \cdot 10^{-6}$
12	217	149	317	150	418	$3.7 \cdot 10^{-6}$	$3.40 \cdot 10^{-6}$
16	215	79.0	171	160	461	$2.0 \cdot 10^{-6}$	$1.65 \cdot 10^{-6}$
20	216	41.9	89.7	180	553	$6.8 \cdot 10^{-7}$	$4.73 \cdot 10^{-7}$
30	231	9.22	18.6	200	647	$2.8 \cdot 10^{-7}$	$1.66 \cdot 10^{-7}$
40	258	2.29	4.12	220	732	$1.3 \cdot 10^{-7}$	$6.82 \cdot 10^{-8}$
50	272	$6.38 \cdot 10^{-1}$	1.09	240	798	$6.6 \cdot 10^{-8}$	$3.11 \cdot 10^{-8}$
60	252	$1.71 \cdot 10^{-1}$	$3.16 \cdot 10^{-1}$	260	853	$3.6 \cdot 10^{-8}$	$1.52 \cdot 10^{-8}$
70	226	$3.74 \cdot 10^{-2}$	$7.69 \cdot 10^{-2}$	280	887	$2.1 \cdot 10^{-8}$	$7.93 \cdot 10^{-9}$
80	214	$7.5 \cdot 10^{-3}$	$1.16 \cdot 10^{-2}$	300	901	$1.3 \cdot 10^{-8}$	$4.42 \cdot 10^{-9}$

The desire to refine these data and obtain characteristics for the higher layers gave rise to a number of investigations. In these the values of the parameters characterizing the atmosphere are given for heights to about 300 km on the basis of various models of its structure. The right hand side of Table 11 gives data calculated on the basis of one of the latest models* in which it is assumed that:

1) pressure, density and temperature should coincide with rocket data to a height of 100 km;

2) molecular oxygen begins to dissociate at a height of 90 km, and at a height of 130 km O_2 molecules are still present in amounts of approximately 25± 5%;

* Sovetskie sputniki i kosmicheskie rakety (Soviet Satellites and Cosmic Rockets).—Izd. AN SSSR. Moskva. 1959.

3) molecular nitrogen (N_2) begins to dissociate at a height of over 220 km;

4) the decrease in the O_2 and N_2 concentration with height is exponential;

5) the kinetic temperature of the exosphere is taken to be constant. At the base of the exosphere (critical level) the temperature should be above 500°K and below 2000°K. The temperature gradient in the layer from 100 to 300 km cannot exceed 5° per kilometer;

6) the density of neutral particles corresponds to ionospheric data and is not less than 10^7 particles per cm^{-3} at the critical level;

7) the influence of diffusive separation on pressure and density to a height of 300 km is small and can be disregarded.

With regard to the temperature in layers above 300 km and in the exosphere, only the tentative data cited above are available; temperatures here, according to various authors, reach very high values (1000–3000°K).

In view of this it is very interesting to consider data obtained by means of artificial satellites. Below we cite results obtained by processing measurements carried out on the 15th orbit (16 May 1958) of the third Soviet satellite (Table 12).

TABLE 12

Structural parameters of atmosphere at altitudes of 225-500 km for 16 May 1958, based on data of third artificial satellite

Height (km)	Molecular weight (g/mole)	N (cm^{-3})	ρ (g/cm^3)	T (°K)	p (mm Hg)
225	21.28	$6.01 \cdot 10^9$	$2.12 \cdot 10^{-13}$	936	$6.25 \cdot 10^{-7}$
250	20.15	$3.3 \cdot 10^9$	$1.1 \cdot 10^{-13}$	958	$3.54 \cdot 10^{-7}$
300	18.50	$1.15 \cdot 10^9$	$3.53 \cdot 10^{-14}$	1048	$1.37 \cdot 10^{-7}$
350	17.47	$4.82 \cdot 10^8$	$1.4 \cdot 10^{-14}$	1185	$6.58 \cdot 10^{-8}$
400	16.84	$2.36 \cdot 10^8$	$6.6 \cdot 10^{-15}$	1373	$3.79 \cdot 10^{-8}$
450	16.43	$1.32 \cdot 10^8$	$3.6 \cdot 10^{-15}$	1614	$2.53 \cdot 10^{-8}$
500	16.16	$8.24 \cdot 10^7$	$2.21 \cdot 10^{-15}$	1953	$1.94 \cdot 10^{-8}$

From the data cited in Table 12 as well as from the results of many other investigations it is clear that the density of air at great altitudes is very small. Like the pressure, it decreases by a factor of about 10^6 at a height of 100 km and 10^{10} at a height of 200 km. The number of molecules per unit volume decreases in roughly the same proportion; as their mean free path also increases (roughly) in this proportion, by a height of about 300 km the mean free path is of the order of kilometers.

Notwithstanding the tentative character of the data one may safely state that temperatures at heights exceeding 300 km are very high and vary quite strongly in time (over the year). It is equally certain that the pressure and density in these layers decrease comparatively slowly with altitude.

It should be mentioned that the values of the air density at high altitudes as determined from observations of the orbits of three Soviet satellites and the American satellites "Explorer-1" and "Vanguard-1" exceed those adopted in almost all theoretical models by nearly one order of magnitude. Thus values of $3.0 \cdot 10^{-13}$, $3.5 \cdot 10^{-13}$ and $4.5 \cdot 10^{-13}$ g/cm^3 were obtained for a height of about 220 km from Soviet satellite data, and values of $9.6 \cdot 10^{-15}$ g/cm^3 at a height of 405 km from American satellite data. The

reason for this has not been fully clarified as yet; it may conceivably lie in pronounced daily variations of the density of the upper atmospheric layers.

§ 6. Air currents in the atmosphere. General circulation of the atmosphere. Jet streams

Air currents in the atmosphere have been studied comparatively fully by aerological methods only to a height of 20–30 km. For higher layers our information is very limited and derived chiefly by indirect methods from observations of the displacement of clouds (mother-of-pearl and noctilucent) and meteor trails, and also from observations of the displacement of irregularities of ionization in ionospheric layers and of moving polar auroras. In recent years a certain amount of data has been obtained from observations involving rockets. However, all these data are still very limited and provide only the roughest idea of the circulation of the atmosphere in the upper layers.

The set of basic forms of large-scale air movement in the lower atmosphere responsible for the transfer of large masses of air in the horizontal and vertical directions is called the general circulation of the atmosphere. Such movements include, firstly, those caused by the difference of temperature between high and low latitudes, next movements due to the temperature difference between oceans and continents (monsoons); important members of the general circulation are the cyclones and anticyclones responsible for the exchange of air between latitudes.

The study of the general circulation of the atmosphere — a crucial and at the same time highly complex question — has always attracted a great deal of attention from investigators, for it is these forms of currents, to which is also linked the transfer of heat and moisture in the atmosphere, that are the main factors controlling weather and the formation of climate. Naturally, therefore, problems of the general circulation are paramount in climatology and synoptic meteorology. In climatology they are studied chiefly on the basis of climatological data on pressure and wind, averaged over considerable periods of time (month, season, year). However, when averaged data of this kind are used many individual peculiarities of atmospheric processes may be erased; as a result they can give a false idea of the character of many atmospheric processes.

In synoptic meteorology the general circulation is studied by means of daily synoptic charts coupled with data of aerological observations. Aside from the indicated methods of investigation, this question has also been treated from the theoretical point of view, an aspect considered in detail in textbooks of dynamic meteorology. The result of this work has been the suggestion of a series of schemes and theoretical models for the general circulation of the atmosphere; however, owing to the great mathematical difficulties involved and the absence of a series of reliable initial data providing an accurate picture of reality, the final answer to the question of creating a theory of the general circulation of the atmosphere still remains remote.

We shall not dwell here on the theory of the problem but will confine ourselves to citing certain fundamental experimental data and briefly exposing present views on a few of its aspects.

The original cause of all atmospheric movements is uneven warming of different parts of the earth's surface and corresponding regions of the atmosphere, resulting in nonuniformity of both temperature and air pressure in a horizontal direction. Pressure differences between points located on the same level surface (the ocean in particular) are the immediate cause of air currents. If the force resulting from the difference in pressure (pressure force) were the only one acting on air particles their motion would take place in the direction of decreasing pressure, in a manner similar to water draining from higher to lower levels. This does not happen, however, since moving air is also influenced by other forces, including the deflecting force of the earth's rotation (Coriolis acceleration), the force of friction and, in curvilinear motion, the centrifugal force. These forces and the question of the appearance of motions in the atmosphere will be considered in greater detail in Chapter 22.

In a number of early schemes suggested for the general circulation only one source of motion (equator-pole temperature difference) is considered. The work performed can then be regarded as a result of the operation of a vast thermal machine with a heater at the equator and cooler at the pole. The term first-order thermal machine has been applied by Academician V. V. Shuleikin to this somewhat arbitrary machine. One of the simplest

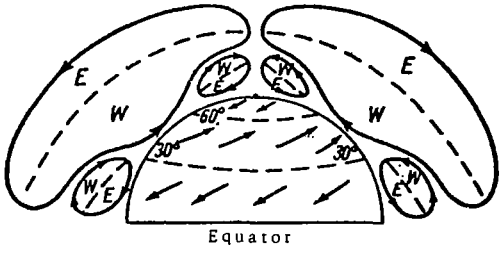


FIGURE 11. Circulation scheme for uniform earth surface

schemes among the many suggested to describe the circulation arising in this case on the uniform surface of the earth when the action of the earth's rotation is taken into account is shown in Figure 11. According to this scheme there are three rings of circulation in each hemisphere. The first lies in the tropics and has an ascending motion near the equator and descending motion near the 30th parallel. The north-easterly trade winds occur in the lower portion of this ring and the westerly anti-trades in the upper portion. The second ring is characterized by a descending motion at the pole and ascending motion near the 60th parallel. North-easterly winds are observed near the ground and westerly winds in the upper section of the ring. The third ring in this scheme — the tropospheric-stratospheric ring — joins the first two to the unified scheme; westerly winds occur in the lower section and easterly winds in the upper section.

This scheme reproduces certain major features of atmospheric circulation but, as research in recent years has shown, it is remote from reality. This is due primarily to the fact that it fails to account for the nonuniformity of the earth's surface, as expressed, firstly, in the presence of oceans

and continents. Indeed, in the cold half of the year, especially in middle and low latitudes, the continents cool off rapidly and severely and act as sinks of heat whereas the oceans, having accumulated thermal energy during the summer, cool off far less rapidly and severely and thus become sources of heat.

The converse relationship obtains in the summer months. The continents, which warm up more rapidly, act as sources of heat, while the oceans act as sinks. This leads to a seasonal alternation of the temperature contrast between land and sea, giving rise to a secondary seasonal circulation between them. The latter may be called (again after Shuleikin) a second-order thermal machine. The circulation resulting from the work of this machine is less powerful than the circulation due to the thermal differences between equator and pole and is largely absorbed by the latter. However, the secondary circulation is manifested in the seasonal alternation of air currents at the surface between ocean and continent (i. e., monsoons).

As the actual circulation of the atmosphere is complicated still further by the constant development of numerous shifting cyclones and anticyclones, it is practically impossible to discern the work of these thermal machines in explicit form. At the very same instant conditions of circulation can be totally different and strongly variable in time along various meridians. As a result it is impossible to express the actual, continually changing circulation over the entire globe in the form of a single scheme, and the attempts of a number of investigators to do so have proved useless in practice. The real character of the circulation in every region must be studied on the basis of whatever one's day-to-day experience reveals about the typical forms, intensity and seasonal pattern of circulation.

Let us now turn to the question of the general character of air currents in the atmosphere. In the troposphere, where they have been studied in greatest detail, air currents in the middle latitudes are westerly; their velocity increases with altitude, reaching a peak at the tropopause (9–12 km). Above the tropopause wind speed decreases with altitude on the average. In summer velocities become minimal at heights of 22–25 km and the wind direction changes, moving gradually to the opposite quarter, i. e., east. In winter the wind continues to be westerly to considerable heights at all latitudes except the equator. This falling-off of wind velocity with altitude above the tropopause is due to the fact that the horizontal temperature gradient changes direction at these altitudes. One of the most important and interesting features of atmospheric circulation is the existence in the upper troposphere of the so-called jet streams. Their study has not only scientific but also great practical significance.

The jet stream is a narrow current of air with a quasi-horizontal axis lying in the upper troposphere, usually at a height of 9–12 km. It is thousands of kilometers long, hundreds of kilometers wide and several (2–4) kilometers thick (vertically). The strongest winds are observed in the central part of the jet where wind velocity is over 30 m/sec and in certain cases reaches 100–150 m/sec or even somewhat more (200 m/sec). Horizontal gradients of wind velocity in this current are very steep: over a distance of 100 km the velocity difference is 10–15 m/sec, increasing sometimes to 25–30 m/sec per 100 km. The vertical gradient of wind velocity is from 1 to 4 m/sec per 100 m of altitude. Usually the jet stream does not lie strictly along the parallel; it is wavy in some areas and sometimes even assumes a meridional direction.

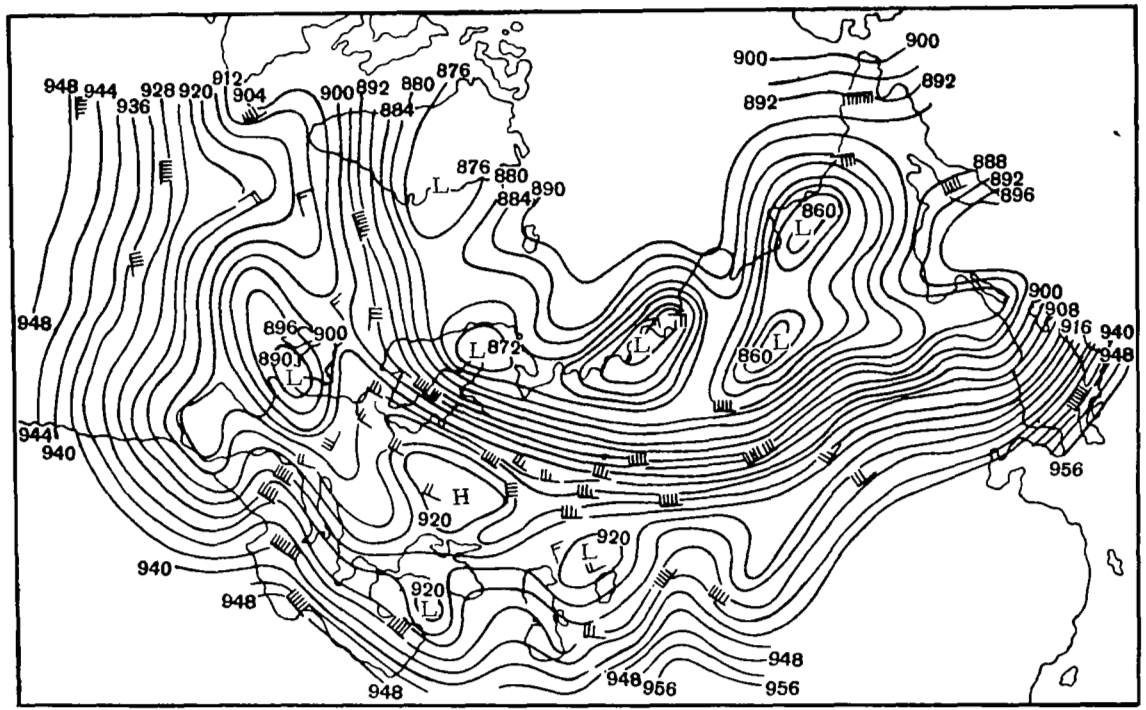


FIGURE 12. Upper-air planetary frontal zones and jet streams

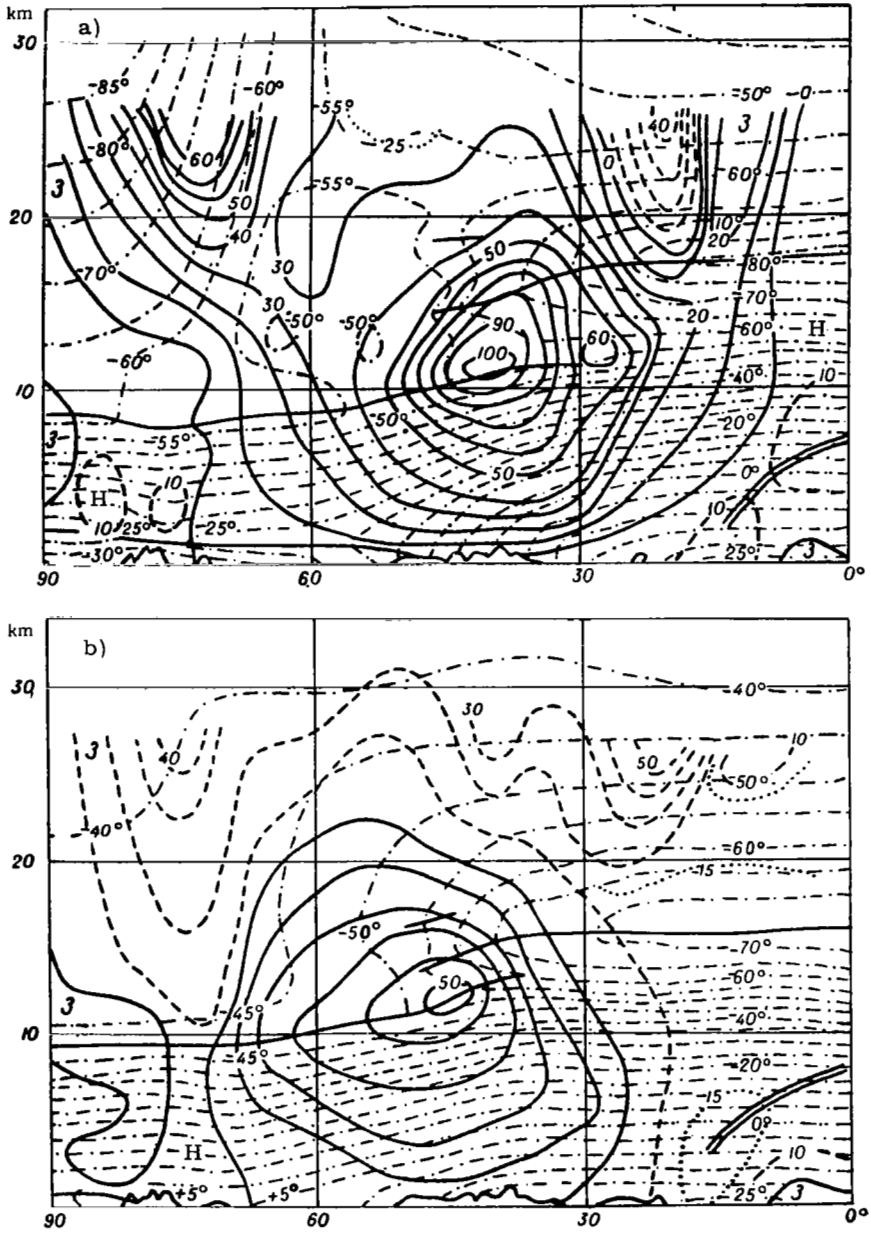


FIGURE 13. Vertical section across the jet stream, along 80th meridian of western longitude

a–January; b–July. Solid lines— isotachs of westerly wind, dotted lines— isotachs of easterly wind, dot-dash lines— isotherms.

Jet streams are observed with varying frequency over practically all regions of the earth; in some areas they display a number of special features. It is important to note the close link between the jet stream and the presence in the atmosphere of the zone of strong horizontal temperature contrasts called the upper-air planetary frontal zone (Figure 12). We also note that in the region of the principal (subtropical) jet stream (Figure 13), lying at a height of 12–12.5 km at latitude 30–35°, the tropopause virtually splits in two: south of the jet stream it occurs at a higher altitude (14–15 km) and is cooler (–60, –70°) while northward it is lower (10–11 km) and warmer (–45, –50°). With the data presently available it is as yet impossible to explain theoretically how and why the jet stream and associated planetary frontal zone were formed.

Air currents at heights exceeding 30 km are characterized on the basis of data obtained by indirect methods.

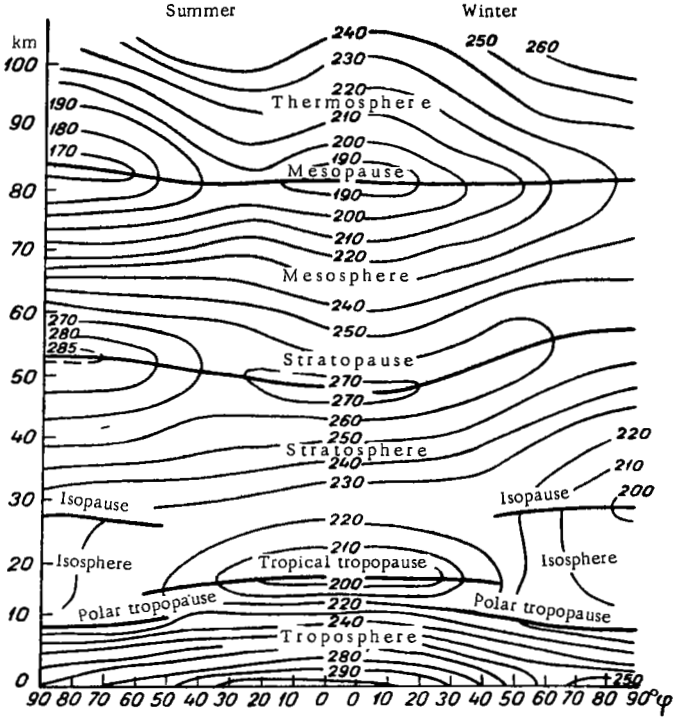


FIGURE 14. Mean meridional cross section of atmosphere (after V.R. Dubentsov)

Aerological observation of the distribution of temperature and pressure makes it possible to rectify our information on air currents in the 35–100 km layer. Figure 14 gives a north-south cross section of the atmosphere in the northern hemisphere in summer and winter (by V. R. Dubentsov). The figure shows that in summer relatively feeble westerly currents predominate in the troposphere from the pole to the tropics while

easterly winds prevail south of the tropics. Above 20 km (in the stratosphere) the wind becomes easterly over the entire hemisphere, reaching a peak velocity at a height of 50–70 km (of the order of 60–70 m/sec); above 80–90 km air currents grade into westerly once more. In winter, as was already mentioned, no such change in wind direction is observed and the wind continues westerly throughout the hemisphere (except the equator) to a height of 40–50 km.

According to sound ranging data, wind velocity increases with altitude and peak velocities are observed at heights of 50–60 km, where mean maximal velocities of about 70–100 m/sec are recorded. The same order of velocity is obtained from observations of meteor trails. Observation of shifting polar auroras gives values as high as 200–300 m/sec for the velocity of the easterly wind at a height of about 100 km. However, it should be borne in mind that in the upper layers air movements can be strongly influenced by a number of factors of no significance for movements in the lower atmosphere. Such factors include electromagnetic forces, which affect charges, ebb- and flow-motions, nonuniform absorption of radiation and many others.

On the basis of the cross sections above and analogous schemes cited by authors outside the USSR (W. Kellogg and Schilling, Pant, et al.) one can draw the general conclusion that the atmosphere consists of three layers differing in the nature of the air currents: from the surface to a height of about 20–25 km in summer and 40–50 km in winter westerly winds predominate; above this level to a height of about 60–70 km lies a considerable layer of easterly winds overlain by a layer of variable winds, and in the region between 80 and 200 km easterly winds are present in low and high latitudes and westerly winds in middle latitudes.

§ 7. Horizontal inhomogeneity of the troposphere. Air masses and fronts

Analysis of the state of the atmosphere over large areas shows that the atmosphere is inhomogeneous in its physical state horizontally as well as vertically. This inhomogeneity is most sharply expressed in the troposphere, where the influence of the nonuniformity of the ground is strongest. Different degrees of warming and cooling of the air over various parts of the ground, differences in humidity, various kinds of motion and turbulence, all produce variations in the properties of the air. However, one will always find fairly large regions of the atmosphere having relatively homogeneous weather conditions and characterized by definite general physical properties. The characteristic dimensions of such regions are in the thousands of kilometers in the horizontal direction and only a few kilometers in the vertical direction (though not infrequently they extend over the entire height of the troposphere). These homogeneous volumes of atmospheric air are called **air masses**. The latter are continuously moving along the surface while retaining their properties over long periods. With the alternation of air masses are connected all very severe aperiodic weather changes at any point (area).

Contrasting air masses border on each other, and between them form the comparatively narrow transitional areas called **frontal zones**.

Along these zones a number of meteorological elements (temperature, humidity, wind) undergo pronounced changes. Frequently these changes are so drastic and the frontal zones so narrow (they generally do not exceed a few hundred kilometers) that one may speak of the surface of discontinuity between two adjacent air masses. In this case it is called the frontal surface or front.

Frontal surfaces lie at a certain small angle to the horizontal surface, and the line of their intersection with the surface of the ground is called frontal line. The sudden, severe change in the meteorological elements occurs as one crosses this line (or, more precisely, narrow belt).

The properties of air masses and fronts are studied and clarified with the help of synoptic charts. This term designates ordinary geographic maps of some sufficiently large territory (sometimes the whole earth) on which have been plotted meteorological or aerological data (fundamental meteorological elements) for some instant in time; these data are based on observations performed at many points.

If one has a chart showing the values of meteorological data at a series of points one can draw up a general summary of the state of the atmosphere over the corresponding territory. By compiling such charts for a number of successive instants it is possible to trace the development of processes unfolding in the atmosphere and covering large areas.

The spatial study of the physical properties of the atmosphere and processes inside it by means of such charts is called the synoptic method. It is widely used in meteorology. Its most important application is in weather forecasting, and thus synoptic charts are often called weather charts.

Experience gained from synoptic charts shows that a constant movement of air masses with different properties takes place in the atmosphere. Air masses which intrude into any given area have different properties depending on the geographic conditions under which they were formed and the surface over which they travelled before reaching the given area.

Consequently, the concept of the so-called source regions of air masses has been introduced. Indeed, if a certain air mass is held up for a long time over, say, the Arctic ice fields it will assume certain properties, over the subtropics other, entirely different properties, and over the hot Sahara sands different properties again.

Accordingly, in the initial stages of investigation into this problem the so-called geographical classification of air masses gained wide currency. In its most general form this classification reduces to the following geographical types of air masses:

1) arctic air (AA), formed in the Arctic basin and adjoining sections of continents (mainly beyond the Arctic circle);

2) middle-latitude air (MA), or polar air (PA);

3) tropical air (TA), formed in tropical and subtropical regions and, in summer, partly also in the southern parts of the middle zone over continents. Each of these types is further subdivided into maritime and continental depending on whether it was formed over oceans or land masses;

4) equatorial air (EA), transitional from one hemisphere to the other.

On the other hand, once it begins to travel away from its source region, every air mass is subjected to the influence of the underlying surface of the regions it crosses. Thus it is perpetually changing its properties. At

any instant the air mass is at a certain stage in its development and its initial physical properties may change fundamentally. This process is called the transformation of air masses. Hence the geographical classification of air masses is insufficient and incomplete.

A detailed study of air masses, especially with the introduction of aerological data, shows that it is physically more correct to classify them by their thermodynamic properties.

The most general classification of air masses contains two categories, warm and cold; a mass is called warm (or cold) if it is warmer (or cooler) than the mass situated next to it. Such a mass is usually warmer (or cooler) than the underlying surface as well. It is then termed stable (or unstable) depending on its characteristic temperature stratification.

As air masses move in the atmosphere the surfaces of discontinuity or fronts travel with them. The direction and speed of a front are determined by the distribution of air currents in its vicinity. Two basic types of fronts are differentiated according to the type of displacement peculiar to the frontal line:

1) warm fronts, fronts moving toward the cold air mass. The warm air usually rises along the slope of cold air which gives way to the advancing warm air;

2) cold fronts, fronts moving toward the warm air mass. In this case the cold air pushes under the warm air, displacing the warmer mass.

Without dwelling on the question of the classification of fronts, which is considered in detail in textbooks of synoptic meteorology, we confine ourselves to noting that a distinction is made between major and secondary fronts. The former cover extensive regions and, taking the geographic classification of air masses as a base, include:

1) arctic fronts, between arctic and polar air;

2) polar fronts, between polar (middle-latitude) and tropical air, and

3) tropical fronts, between tropical and equatorial air.

Secondary fronts cover much smaller areas and usually separate various parts of one air mass belonging to a single type.

The vertical and horizontal cross section of a frontal zone (frontal surface) are represented schematically in Figure 15.

Surfaces of discontinuity between air masses (frontal surfaces) have certain specific attributes. The main one is that meteorological elements undergo a sharp change (discontinuity). Changes of this kind are observed in the march of temperature, density, humidity, wind velocity and that component of the pressure gradient which is normal to the surface of discontinuity. The pressure, however, changes continuously on crossing the surface of discontinuity. The most substantial discontinuity is in the temperature, which can change by several degrees (sometimes up to 10–15°) as one moves from the warm to the cold mass.

As we noted earlier, the frontal surface always lies at a small angle to the horizontal. This is a necessary consequence of the fact that under the influence of the deflective force of the earth's rotation two air masses at different temperatures can be in equilibrium only if they are moving relative to each other while separated by an inclined surface of discontinuity. In the simplest case, the theory shows that the angle of inclination α of the surface of discontinuity can be expressed in terms of the width of the

frontal zone AB and its vertical height BC (Figure 15) by the relation

$$\operatorname{tg}\alpha = \frac{BC}{AB} = \frac{2\omega \sin\varphi}{g} \frac{T_1 v_2 - T_2 v_1}{T_1 - T_2},$$

where φ is the latitude, g the acceleration of gravity, ω the angular velocity of rotation of the earth, and T_1, v_1 and T_2, v_2 the temperature and wind speed in the warm (subscript 1) and cold (subscript 2) mass respectively.

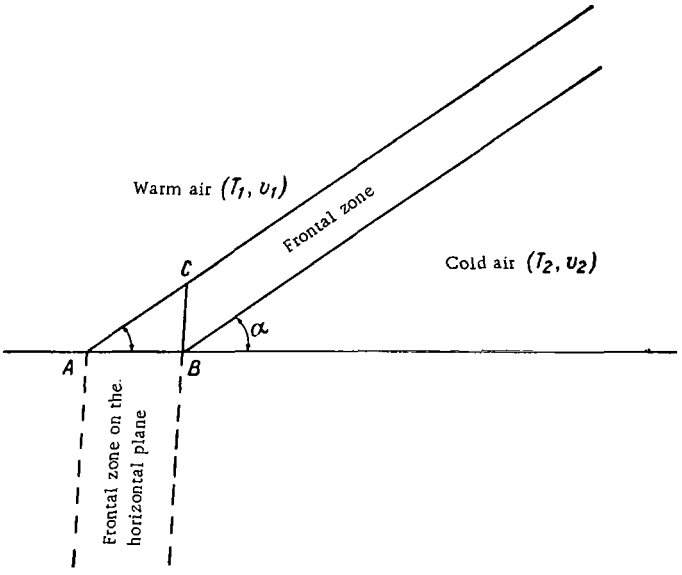


FIGURE 15. Schematic representation of frontal zone

From this relation it is obvious that the angle of inclination α depends on latitude, increasing with increasing latitude. Further, it is a function of the discontinuity in temperature and wind speed and also depends on the values of the latter quantities. With regard to the numerical values of α, we note that in middle latitudes for $T_2 = 273°$, $\Delta T = T_1 - T_2 = 2.5°$ and $v_1 = 0$ we find the values $\operatorname{tg}\alpha = \frac{1}{177}$ for $v_2 = 5$ m/sec and $\operatorname{tg}\alpha = \frac{1}{88}$ for $v_2 = 10$ m/sec.

Usually the angle of inclination is really very small: $\operatorname{tg}\alpha \approx \frac{1}{100} - \frac{1}{200}$, i.e., α is only a few tenths of a degree.

Since different air masses have different properties and highly complex processes develop in the atmosphere near frontal surfaces, their displacement, formation and dissipation are accompanied by changes in weather conditions at all points on the ground. The displacement of fronts brings characteristic changes in the clouds, precipitation, and changes in the temperature, humidity and other meteorological elements by which weather is characterized. Hence the detailed study of fronts, their properties, motions, and conditions of formation and dissipation constitutes one of the principal aims of synoptic meteorology. However, a consideration of these problems would take us beyond the scope of the present textbook. The fundamental concepts and definitions introduced above are all that need concern us.

Chapter 5

DISTRIBUTION OF PRESSURE IN THE ATMOSPHERE

§ 1. Hydrostatic equation of the atmosphere

Let us consider a vertical column of air of unit cross section in the atmosphere. We will assume that: 1) the air is at rest relative to the earth's surface; 2) it can be regarded as an ideal gas; 3) the composition of the air does not change with height. Then for the air to be in equilibrium at the height z it is necessary that its buoyancy p balance the weight Q of the column above the given level, i. e., that

$$Q = p. \tag{1}$$

From this equation it follows that as one moves upward the buoyancy p of air decreases as a consequence of the decrease of the weight Q. Thus when we measure the buoyancy of air we are thereby measuring the force exerted on a unit surface by the weight of the air column standing vertically above the given level. This force we define as the atmospheric pressure, denoting it in the same way as the buoyancy by p.

Strictly speaking, in the case where the air is in motion the relation (1) is not fulfilled. However, detailed analysis of this question shows that in reality air motions and accelerations in the atmosphere are so small that their influence can be disregarded in practice. Only in the case of very high velocities and especially strong vertical accelerations can one detect a certain insignificant influence of the latter on the pressure.

Let us consider the case when vertical air movements are absent. We take a column of unit cross section at any height in the atmosphere. Let the pressure at its base be p and at its top $p-dp$. It is obvious that in the absence of a pressure difference in the horizontal direction the pressure drop $-dp$, from equation (1), is given by the weight of the column of air. If ρ is the density of the air at the given height z, and g the acceleration of gravity, then

$$-dp = \rho g dz. \tag{2}$$

This relation, familiar to all from physics textbooks, relates pressure and density to height for an ideal gas subject to the influence of gravity. It is valid under the conditions of static equilibrium indicated above and is usually called the fundamental equation of hydrostatics of the atmosphere. From this equation it follows directly that the fall of pressure with height is directly proportional to the density of air.

§ 2. Barometric formulas

Starting from the equation of hydrostatics (2) it is easy to obtain the general law of variation of pressure and density with height. Indeed, if we make use of the equation of state (Mendeleev-Clapeyron equation)

$$\rho = p\frac{\mu}{R^*T} = \frac{p}{RT}$$

and introduce the value of ρ into equation (2), the latter assumes the form

$$-dp = p\frac{\mu g}{R^*T}\,dz = \frac{pg}{RT}\,dz. \tag{3}$$

From this, separating the variables and integrating between the limits $z=0$, where the pressure is p_0, and $z=z$, where $p=p_z$, we obtain the so-called barometric formula in its general form

$$p_z = p_0 e^{-\int_0^z \frac{g\,dz}{RT}} = p_0 e^{-\int_0^z \frac{g\mu}{R^*T}\,dz}, \tag{4}$$

which gives the law of variation of pressure with altitude.

Further, since $\rho_z = \frac{p_z}{RT_z}$ and $\rho_0 = \frac{p_0}{RT_0}$, whence

$$\frac{p_z}{p_0} = \frac{\rho_z T_z}{\rho_0 T_0},$$

we obtain the general law of variation of density with altitude

$$\rho_z = \rho_0 \frac{T_0}{T_z} e^{-\int_0^z \frac{g\mu}{R^*T}\,dz}, \tag{4'}$$

and since $\rho = nm$, where n is the number of molecules per cubic centimeter and m the mean mass of a molecule, we also have

$$n_z = n_0 \frac{T_0}{T_z} e^{-\int_0^z \frac{g\mu}{R^*T}\,dz}, \tag{4''}$$

which gives us the variation with height of the number of molecules per unit volume.

As one can see from the latter formulas, the pressure and density variation of any gas with height in the atmosphere depends on the type of gas (molecular weight μ), the variation of the temperature T and the acceleration of gravity g.

Owing to the fact that the law of variation of air temperature with height cannot generally be expressed by a simple analytic formula, the integration of equation (4) cannot be carried out in a general form. It can be performed only approximately, or for particular cases with specified height distribution of the temperature. In this case it is usual to regard the composition of the air as invariable ($\mu = \text{const}$) and disregard the variation of g with height, assuming that $g(z) = \text{const}$. This, of course, introduces a certain element of inaccuracy and is admissible to altitudes of about 80–100 km.

In meteorology the following particular cases are usually considered:

1) the density ρ of the atmosphere does not vary with height; this is the so-called homogeneous atmosphere;

2) the temperature remains constant with height in the atmosphere (T = const); this is the isothermal atmosphere;

3) air temperature in the atmosphere decreases with altitude according to the linear law $T_z = T_0 - \gamma z$, where $\gamma = -\frac{\partial T}{\partial z}$ is the vertical temperature gradient; this is the so-called polytropic atmosphere.

While these cases are inapplicable for the atmosphere as a whole, it is nevertheless meaningful to single them out because, up to the altitudes investigated, the atmosphere can be split into a series of layers each corresponding more or less to one of the indicated conditions. Let us consider the pressure variation under these conditions.

1. Homogeneous atmosphere. Here $\rho(z) = \text{const} = \rho_0$, and since $g = \text{const} = g_0$, integration of equation (2) gives

$$p_z = p_0 - g\rho_0 z. \tag{5}$$

Consequently, the pressure in such a homogeneous (arbitrary) atmosphere decreases linearly and becomes zero at the height

$$z = \frac{p_0}{\rho_0 g_0} = \frac{R^* T}{\mu g_0} = \frac{RT}{g_0} = H. \tag{6}$$

This height, which is usually denoted by H, is called the height of the homogeneous atmosphere or scale height. Its numerical value for, say, dry air (μ = = 28.966), with $g_0 = 9.8$ m/sec^2 and $T = T_0 = 273$°K, is $H_0 = 7990$ m ≈ 8000 m.

For a temperature T different from $T_0 = 273$°K (0°C), as is easy to show,

$$H = \frac{RT}{g_0} = \frac{RT_0 T}{g_0 T_0} = H_0 \frac{T}{T_0} = H_0 (1 + \alpha t). \tag{6'}$$

The values of H_0 for some of the gases present in atmospheric air is given in Table 13.

TABLE 13

Height of homogeneous atmosphere (H_0) for various gases

Gas	Nitrogen N_2	Oxygen O_2	Argon A	Carbon dioxide CO_2	Hydrogen H_2	Water vapor H_2O	Air
H_0 (m)	8860	7230	5800	5225	114 980	12 830	7990

The variation of temperature with altitude in the homogeneous atmosphere can be found from the equation of state, from which it follows that $T = \frac{p}{R\rho}$. Differentiating this equation with respect to z under the condition ρ = const and taking the fundamental equation of hydrostatics (2) into account, we obtain

$$\frac{dT}{dz} = \frac{1}{\rho R}\frac{dp}{dz} = -\frac{g}{R} = -3.42°/100\text{m}. \tag{7}$$

This is the value of the vertical temperature gradient for the homogeneous atmosphere.

The concept of the height of the homogeneous atmosphere H is arbitrary, but this quantity is meaningful as an auxiliary constant and is frequently used in many calculations. With the help of this quantity it is possible to modify the equation of hydrostatics (2). Indeed, from (3), taking (6) into

account, we obtain

$$-\frac{dp}{p}=\frac{g}{RT}dz=\frac{1}{H}dz, \tag{8}$$

from which it is evident that $H=\frac{RT}{g}$ can be chosen as a unit for measuring altitudes.

2. Isothermal atmosphere. If $T=$ const then, integrating the equation of hydrostatics (3) between the limits (p_0, p_z) and $(0, z)$, we find that

$$\ln\frac{Tp_z}{p_0}=-\frac{g_0z\mu}{R^*T}=-\frac{g_0z}{RT}, \tag{9}$$

or

$$p_z=p_0e^{-\frac{g_0\mu}{R^*T}z}=p_0e^{-\frac{g_0z}{RT}}. \tag{9'}$$

Since according to the preceding $\frac{\mu g_0}{R^*T}=\frac{1}{H}$

$$p_z=p_0e^{-\frac{z}{H}}, \tag{9''}$$

i.e., pressure in the isothermal atmosphere decreases exponentially with altitude.

From this it is obvious that the isothermal atmosphere has no upper boundary and extends to infinity. The pressure inside it decreases only by the factor e at the height $z=H$, where pressure in the homogeneous atmosphere would be zero.

From (9'') it is seen that the height z at which the pressure equals p_z in the isothermal atmosphere can be expressed by the formula

$$z=H\ln\frac{p_0}{p_z} \tag{10}$$

or, taking into account (6') and passing over to common logarithms,

$$z=H_0(1+\alpha t)2.3\log\frac{p_0}{p_z}.$$

Taking $H_0=8000$m, the last equation can be rewritten in the form

$$z=18\,400(1+\alpha t)\log\frac{p_0}{p_z}. \tag{10'}$$

From the above it is easy to find, in particular, that pressure in the isothermal atmosphere should decrease (for $t=0°$) 10 times at a height of 18.4 km and 1000 times at a height of about 37 km. In reality this height is somewhat less, since the mean temperature of this layer is less than zero.

3. Polytropic atmosphere. In this atmosphere it is assumed that temperature changes linearly with height according to

$$T=T_0-\gamma z,$$

where $\gamma=-\frac{\partial T}{\partial z}$ is the vertical temperature gradient.

For such an atmosphere the equation of hydrostatics (3) has the form

$$\frac{dp}{p}=-\frac{gdz}{RT}=-\frac{gdz}{R[T_0-\gamma z]}=\frac{g}{R\gamma}\frac{dT}{T}. \tag{11}$$

Its integration between the limits p_0 to p_z and T_0 to T_z for $g = \text{const}$ gives

$$\ln \frac{p_z}{p_0} = \frac{g}{R\gamma} \ln \frac{T_z}{T_0},$$

or

$$\frac{p_z}{p_0} = \left[\frac{T_z}{T_0}\right]^{\frac{g}{R\gamma}} = \left[1 - \frac{\gamma z}{T_0}\right]^{\frac{g}{R\gamma}}, \tag{12}$$

which expresses the variation of pressure with height in the polytropic atmosphere.

The height z at which the pressure equals p_z in the polytropic atmosphere can be found from (12); thus

$$z = \frac{T_0}{\gamma}\left[1 - \left(\frac{p_z}{p_0}\right)^{\frac{\gamma R}{g}}\right], \tag{13}$$

i.e., the height depends on the numerical value of the vertical temperature gradient.

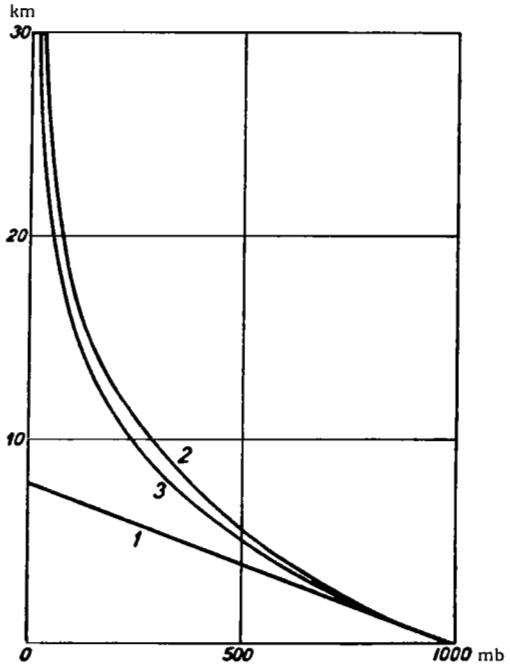


FIGURE 16. Pressure variation in homogeneous (1), isothermal (2) and polytropic (3) atmosphere

The upper boundary of the polytropic atmosphere, i.e., the height at which $p_z = 0$, can be obtained from (13)

$$Z = \frac{T_0}{\gamma}. \tag{14}$$

For example, for $T_0 = 273°$ and $\gamma = 6°/\text{km}$ $Z = 45\,\text{km}$, while for $\gamma \to 0$ $Z \to \infty$.

It is easy to show that the homogeneous atmosphere is a particular case of the polytropic atmosphere. Indeed, in the homogeneous atmosphere $\gamma = \frac{g}{R} = 3.42°/100\,\text{m}$, and for $T_0 = 273°$ we find from (14) that

$$Z = \frac{T_0}{\gamma} \approx 8000 \text{ m}. \tag{15}$$

If one compares the vertical pressure variation in the three particular cases considered (Figure 16) one sees that pressure falls off most rapidly in the homogeneous atmosphere, while in the isothermal atmosphere it decreases somewhat more slowly than in the polytropic atmosphere.

A certain arbitrary standard atmosphere characterized by a near-average vertical distribution of the meteorological elements is introduced in various kinds of practical computations (chiefly with application to aviation). For this atmosphere it is assumed that temperature decreases linearly with a gradient of 0.0065°/m to a height of 11 km, while from 11 km to 30 km it remains constant and equal to 216°K (or −56.5°C). Thus the standard atmosphere is polytropic to a height of 11 km and isothermal above this level, so that the pressure drop can be expressed by the corresponding formulas.

§ 3. Complete barometric formula (Laplace formula)

Having considered the particular cases we now turn to the barometric formula in general form (4). We write it in the form

$$p_2 = p_1 e^{-\int_{z_1}^{z_2} \frac{g}{RT} dz}, \tag{16}$$

where p_1 and p_2 are the pressure at the levels z_1 and z_2 respectively.

In the preceding section where particular cases were considered, we specified the law of variation of temperature with height. However, it is possible to divide the atmosphere into a series of comparatively thin layers and introduce for each of these the so-called mean barometric temperature given by the rleation

$$\bar{T}_{\text{bar}} = \frac{z_2 - z_1}{\int_{z_1}^{z_2} \frac{dz}{T}}. \tag{17}$$

Then instead of (16) we can write

$$p_2 = p_1 e^{-\frac{g}{R} \frac{z_2 - z_1}{\bar{T}_{\text{bar}}}}, \tag{18}$$

and the difference between the two heights $(z_2 - z_1)$ will be given by the formula

$$z_2 - z_1 = \frac{R}{g} \bar{T}_{\text{bar}} \ln \frac{p_1}{p_2}. \tag{19}$$

The meaning of the quantity $\bar{T}_{\text{bar}}$ introduced here is obvious. It is that temperature of the fictitious isothermal layer for which its thickness and

the pressure difference between its boundaries are equal to the thickness and pressure difference between the boundaries of the actual atmospheric layer. In practice, however, the mean barometric temperature $\overline{T}_{bar}$ is rarely calculated and the mean temperature $\overline{T}=\frac{T_1+T_2}{2}$ is usually used instead. The latter is somewhat greater than the mean barometric temperature $(\overline{T}>\overline{T}_{bar})$.

The error introduced in the determination of the pressure by substituting the mean temperature for the mean barometric temperature is small. It can exceed one millibar, e.g., the accuracy of pressure measurements in aerological observations, only for thick layers ($(z_2-z_1)>5$ km) and large values of the vertical temperature gradient ($\gamma = 1°/100$ m).

Let us introduce the mean temperature $\overline{T}$ of the layer z_2-z_1 into formula (19). Then after integration we arrive at the expression

$$z_2-z_1=\frac{R}{g}\overline{T}\ln\frac{p_1}{p_2}. \tag{20}$$

In all these estimates no allowance was made for the fact that real air is always moist and that the acceleration of gravity changes with latitude and altitude. To account for the influence of the water vapor contained in air, as was shown in § 4 of Chapter 2, the real temperature in the equation of state should be replaced by the virtual temperature

$$T_v=T\left[1+\beta\left(\frac{e}{p}\right)\right]=T_0(1+\alpha t)\left[1+\beta\left(\frac{e}{p}\right)\right],$$

where $\beta = 0.378$.

Then all the relations derived earlier become valid for moist air as well and instead of (20) we write

$$z_2-z_1=\frac{R}{g}\overline{T}_v\ln\frac{p_1}{p_2}=\frac{R}{g}T_0(1+\alpha\overline{t})\left[1+\beta\left(\frac{\overline{e}}{p}\right)\right]\ln\frac{p_1}{p_2}, \tag{21}$$

where $T_0 = 273°$K, and $\overline{t}$ and $\left(\frac{\overline{e}}{p}\right)$ are the mean values of these quantities for the layer in question.

It is well known that the variation of the acceleration of gravity with altitude and latitude is given by the relation

$$g=g_{0.45}(1-a\cos 2\varphi)(1-bh).$$

Here φ is the latitude, h the height above sea level, $g_{0.45}=980.6$ cm/sec^2 (acceleration of gravity at sea level for latitude $\varphi = 45°$), $a = 0.00264$ and $b=3.14\cdot 10^{-7}$ m^{-1}.

To introduce the correction for the variation of the acceleration of gravity, it is necessary to replace g by

$$g=g_{0.45}(1-a\cos 2\varphi)(1-b\overline{h}), \tag{22}$$

where $\overline{h}=\frac{z_1+z_2}{2}$ is the mean height of the layer.

Then instead of (21) we obtain

$$z_2-z_1=\frac{RT_0(1+\alpha\overline{t})\left[1+\beta\left(\frac{\overline{e}}{p}\right)\right]}{g_{0.45}(1-a\cos 2\varphi)(1-b\overline{h})}\ln\frac{p_1}{p_2} \tag{23}$$

or, taking into account the smallness of $a\cos 2\varphi$ and $b\bar{h}$,

$$z_2 - z_1 = \frac{RT_0}{g_{0.45}}(1+\alpha\bar{t})\left(1+\beta\left(\frac{\bar{e}}{p}\right)\right)(1+a\cos 2\varphi)\times(1+b\bar{h})\ln\frac{p_1}{p_2}. \quad (24)$$

The formula obtained is called the Laplace formula. Replacing the natural by the common logarithms and introducing the numerical values of the quantities which appear in it, we obtain the Laplace formula in final form

$$z_2 - z_1 = 18\,400\,(1+0.00366\bar{t})\left[1+0.378\left(\frac{\bar{e}}{p}\right)\right]\times$$

$$\times(1+0.00264\cos 2\varphi)\,(1+3.14\cdot 10^{-7}\bar{h})\log\frac{p_1}{p_2}. \quad (25)$$

This formula makes it possible to determine exactly the altitude difference $z_2 - z_1$ between two points at which the pressure is p_2 and p_1 respectively. In practice, however, it is rarely used and simpler formulas are employed. In particular, if we regard the air as dry $\left(\frac{e}{p}=0\right)$ and disregard the latitude and altitude dependence of the acceleration of gravity, we obtain formula (10')

$$z_2 - z_1 = 18\,400\,(1+0.00366\bar{t})\log\frac{p_1}{p_2}. \quad (26)$$

§ 4. Simplified barometric formula. The barometric step

In many cases where the variation of the pressure in layers of limited thickness is considered, a simplified barometric formula (Babinet's formula) is employed:

$$z_2 - z_1 = 2H_0(1+\alpha\bar{t}_v)\frac{p_1-p_2}{p_1+p_2}, \quad (27)$$

where

$$2H_0 = 2\frac{273R}{g} \approx 16\,000 \text{ m}. \quad (27')$$

This is Babinet's formula which, however, can be used only for comparatively thin layers. It can be derived easily from equation (9'), which we will write in the form

$$p_2 = p_1 e^{-\frac{z_2-z_1}{H}} \quad (28)$$

Expansion in series of the quantity $e^{-\frac{z_2-z_1}{H}}$ makes it possible to express (28) approximately as

$$p_2 = p_1\left(1-\frac{z_2-z_1}{H}\right),$$

whence we find that

$$z_2 - z_1 = H\frac{p_1-p_2}{p_1}.$$

Recalling that $H=H_0(1+\alpha t)$ and introducing the mean virtual temperature, we obtain

$$z_2 - z_1 = H_0(1+\alpha\bar{t}_v)\frac{p_1-p_2}{p_1}, \quad (28')$$

or, numerically,

$$z_2 - z_1 = 8000\,(1 + \alpha\bar{t}_v)\,\frac{p_1 - p_2}{p_1}.$$

If high accuracy is not required the vertical variation of the pressure can be characterized by the so-called barometric step, which is the height (in meters) of ascent or descent required for a pressure change of one millibar. From the fundamental equation of hydrostatics (2), written in finite differences, we have for the barometric step

$$-\frac{\Delta z}{\Delta p} = \frac{1}{g\rho} = \frac{RT}{gp} = \frac{8000\,(1+\alpha t)}{p}, \tag{29}$$

i.e., the value of the barometric step is inversely proportional to the air density ρ and, consequently, increases with altitude. For the same pressure p the barometric step is greater in warm than in cool air. The expression for the barometric step is also easily obtained from the simplified barometric formula, from which we immediately recover (29).

From (29) we see that, for $p = 1000$ mb and $t = 0°$, $\frac{\Delta z}{\Delta p} = 8$ m/mb. Values of the barometric step for other p and t are given in Table 14. The table illustrates the indicated dependence of the barometric step on height (pressure) and temperature.

TABLE 14

Values of the barometric step (m/mb)

Pressure (mb)	Temperature (degrees)				
	-40	-20	0	+20	+40
1000	6.7	7.4	8.0	8.6	9.3
500	13.4	14.7	16.0	17.3	18.6
100	67.2	73.6	80.0	86.4	92.8

The inverse value of the barometric step gives the pressure change on ascent by the unit height, i.e., the vertical component of the pressure gradient, or vertical pressure gradient. We denote it by G_z and, since pressure decreases with altitude, we can write

$$G_z = -\frac{\partial p}{\partial z}. \tag{30}$$

Bearing in mind that

$$\frac{\partial p}{\partial z} = -g\rho = -\frac{g}{R}\frac{p}{T},$$

we find

$$G_z = \frac{981}{2.87 \cdot 10^6}\frac{p}{T} = 3.42\frac{p}{T}\ \text{mb/100 m}. \tag{30'}$$

Using (30') one can easily obtain the values of G_z for different p and t. They are listed in Table 15.

As one can see from these data, the vertical pressure gradient is greatest in the lower troposphere, especially at low temperatures: it decreases rapidly with altitude. We note that vertical pressure gradients

TABLE 15

Values of the vertical pressure gradient (mb/100m) for various p and t

Temperature (degrees)	Pressure p (mb)		
	1000	500	100
-40	14.68	7.34	1.47
0	12.52	6.26	1.25
+40	10.92	—	—

are many times larger than horizontal ones. As we will see later on, the mean horizontal pressure gradient is about 1 mb/100 km.

§ 5. Applications of the barometric formula

The barometric formula relates three fundamental quantities: the pressure p, the virtual temperature T_v and the height z. If two of these are determined from observations the third can be found by using the relations derived above. In practical work the following problems are encountered.

1) **Calculation of the vertical distribution of pressure.** The problem consists of determining the magnitude of the pressure p_2 at a certain level z_2 from the specified value p_1 at the level z_1 and mean value $\bar{t}_v$ in the layer $z_2 - z_1$. In these calculations the formulas for the isothermal atmosphere are usually used.

2) **Barometric levelling.** This is used when the problem is to determine the height difference $z_2 - z_1$ between two points from the values of the pressures p_2, p_1 and virtual temperatures $(t_v)_2$ and $(t_v)_1$ at these points. Where high accuracy is necessary calculations should be performed with the Laplace formula. In practice, however, calculations are frequently carried out with the formula for the isothermal atmosphere, using it successively for thin layers (1–2 km). To each of these formula (10') is applied in the form

$$\Delta z = \Delta z_0 \alpha \bar{t}_v, \tag{31}$$

where $\Delta z_0 = 18\,400 \log \frac{p_1}{p_2}$ is the thickness of the layer in which pressure should change from p_1 to p_2 for $\bar{t}_v = 0°$.

In practice the so-called hypsometric tables and nomograms are usually used for barometric levelling.

3) Reduction of pressure to sea level. p_2, $(t_v)_2$, z_2, z_1 and $(t_v)_1$ are known. It is required to find p_1. This question is of frequent occurrence in meteorology, e.g., in reducing the pressure p_2 observed at a certain station at the height z_2 to the pressure p_1 at sea level. In practice this is carried out by reference to standard tables.

4) Determination of the mean temperature $\overline{t}_v$ of the layer. In this case p_1 and p_2 at the heights z_1 and z_2 are known and one is required to find $(\overline{t}_v)$; this is done by means of the formulas given above.

§ 6. Applicability of the barometric formula to the upper atmospheric layers

Obviously, when using the barometric formula to calculate the pressure at high altitudes in the atmosphere one cannot regard the acceleration of gravity as constant but must account for its vertical variation, which is given by the relation

$$g_h = g_0 \left(\frac{a}{a+h}\right)^2,$$

where a is the earth's radius.

Bearing this in mind we should write the barometric formula in general form as follows:

$$-dp = \frac{p}{RT} g_0 \left(\frac{a}{a+h}\right)^2 dz,$$

whence we obtain

$$p_h = p_0 e^{-\frac{g_0}{RT}\left(\frac{a}{a+h}\right)h}. \qquad (32)$$

From this expression one sees that the pressure (and also the density) decreases more slowly with altitude than for g = const. Furthermore, this difference becomes perceptible at a level of the order of 100 km and increases with height. This can be seen from Table 16, which cites the values of the ratio of the pressure $\overline{p}_h$, calculated with allowance for the decrease of g with height, to the pressure p_h, calculated without such an allowance.

TABLE 16

Values of the ratio $\frac{\overline{p}_h}{p_h}$

	h (km)				
	100	200	300	400	500
T = 300°K	1.19	2.00	4.63	14.63	62.8
T = 900°K	1.06	1.26	1.67	2.45	3.98

Aside from the above it is also necessary to consider that the air temperature in the upper atmospheric layers is a highly complex function of the altitude and that this function has not yet been studied with sufficient accuracy.

Finally, the question arises to what extent the concept of gas pressure is applicable to the upper layers of the atmosphere, where rarefaction is very high and the number of molecules is very small. As collisions between molecules are very infrequent here the mean free path of the molecules is

comparable to the height of the homogeneous atmosphere. However, upon examining this question one is led to conclude that if by temperature is understood the gas-kinetic temperature as a measure of the mean energy of thermal motion of the gas molecules one can use the barometric formula even for highly rarefied gas right up to heights of the order of 1000 km (but not higher)

§ 7. Mass of the atmosphere and its vertical distribution

For the mass of air in a column of unit cross section and height dz we can write

$$dm = \rho dz,$$

(ρ being the density of air).

Consequently, for the entire column of the atmosphere

$$m_0 = \int_0^\infty \rho dz. \tag{33}$$

The total mass of the atmosphere will be

$$M = m_0 s = s \int_0^z \rho dz, \tag{34}$$

where s is the surface of the earth, and

$$\rho = \rho_0 \frac{T_0}{T} e^{-\int_0^\infty \frac{g}{RT} dz} \tag{35}$$

Insertion of (35) into (33) gives an expression from which it is possible to find m_0. However, to do this one must know the distribution of temperature throughout the atmosphere and, of course, allow for the variation of g with height.

By an approximate computation, assuming that $T = T_0$ and does not change with height while g = const, we obtain

$$m_0 = \rho_0 \int_0^\infty e^{-\frac{z}{H}} dz = H\rho_0 \text{ g/cm}^2$$

Hence, assuming that $\rho_0 = 1.27 \cdot 10^{-3}$ and $H = 7991$ m, we find for the total mass M

$$M = sm_0 = 4\pi a^2 H\rho_0 = 5{,}3 \cdot 10^{15} \text{ tons}$$

where $a = 6.37 \cdot 10^8$ cm is the radius of the earth.

Since the mass of the entire earth is estimated to be roughly $6.0 \cdot 10^{21}$ tons it follows that the mass of the atmosphere is 10^6 times smaller than that of the earth.

With the same assumption of an isothermal atmosphere, let us determine the vertical distribution of mass in the atmosphere. Since $dm = \rho dz$, we find that the mass of the atmosphere to the height z is given by

$$m_z = \int_0^z \rho dz = H\rho \left(1 - e^{-\frac{z}{H}}\right) = m_0 \left(1 - e^{-\frac{z}{H}}\right), \tag{36}$$

or

$$\frac{m_z}{m_0} = 1 - e^{-\frac{z}{H}} = 1 - e^{-\frac{z}{8000}}. \qquad (36')$$

Calculations show that in such an atmosphere about 63 % of the entire mass of the atmosphere would be contained within the height $z = 8$ km. Analogously, we find that, assuming an isothermal atmosphere with $t = 0°$, 50 % of its entire mass will be contained within $z = 5.5$ km, 90 % within $z = 18.4$ km and 99 % within $z = 36$ km.

In reality, owing to the decrease of temperature with altitude the numbers are somewhat (though insignificantly) different; the picture, however, remains the same.

§ 8. Factors influencing the variation of pressure

Let us turn to the formula

$$p = p_0 e^{\frac{-gz}{RT_v}}.$$

Taking its logarithm and then differentiating, we obtain

$$\frac{dp}{p} = \frac{dp_0}{p_0} + \frac{gz}{R}\frac{d\overline{T}_v}{\overline{T}_v^2}. \qquad (37)$$

Formula (37) relates the pressure change at the upper and lower levels to the mean temperature $\overline{T}_v$ and its changes in the layer bounded by these levels. Let us consider the following cases.

1. The pressure at the initial level (in particular, the ground) is constant, i.e., $p_0 = \text{const}$, or $dp_0 = 0$. Then

$$dp = p\frac{gz}{R}\frac{dT_v}{\overline{T}_v^2}. \qquad (38)$$

From this it is evident that if $dT > 0$ then $dp > 0$, i.e., for constant surface pressure the pressure p at a definite level z increases as the mean temperature of the air column increases, and vice versa.

2. The pressure at the level z remains constant, i.e., $p = \text{const}$, or $dp = 0$. Then

$$dp_0 = -p_0\frac{gz}{R}\frac{d\overline{T}_v}{\overline{T}_v^2}, \qquad (39)$$

i.e., for constant pressure at a certain height the pressure will decrease below it as the mean virtual temperature of the layer increases. The greater dT_v and the greater the thickness of the layer z, the greater this pressure reduction. It is greater for smaller values of $\overline{T}_v$. Accordingly, if air is cooled, say, by contact with a cold underlying surface, the pressure p_0 increases. When air is warmed over warmer surfaces the pressure should drop ($dp_0 < 0$). Variations of the temperature of the air column due to other causes have the same effect.

3. The temperature of the layer remains constant, i.e., $\overline{T}_v = \text{const}$ and,

therefore, $d\bar{T}_v = 0$. Then

$$\frac{dp}{p} = \frac{dp_0}{p_0}, \tag{40}$$

i.e., a relative change in the pressure at a certain height brought about by any cause is equal to a similar change in pressure at the lower level. In absolute value dp is $\frac{p}{p_0}$ times smaller than dp_0, i.e., a small variation of the pressure at a certain height is accompanied by considerable changes in the pressure below. Thus if $p_0 = 1000$ mb and $p = 200$ mb, a 1 mb change in p will cause a 5 mb change in p_0.

These relations are all important because they make it possible to estimate changes in the pressure at any altitude, and also in the mean virtual temperature, from the character of pressure changes observable at the surface. It should be remarked, however, that at any time pressure variations both at the surface and at any height are determined by the combined influence of various factors acting throughout the atmospheric column, and therefore the overall picture of pressure changes is more complex.

§ 9. Isobaric surfaces and isobars. Pressure regions

By drawing surfaces through points in the atmosphere having the same pressure we obtain isobaric surfaces. These are generally fairly complex in form.

The intersection of isobaric surfaces with the sea-level surface or with the surface of any other level z produces curved lines, the so-called isobars of this level (equal pressure lines). Obviously, isobars can neither come into contact nor intersect each other, and each isobar is characterized by a definite value of the pressure. The intersection of isobaric surfaces with the vertical plane (e.g., plane of the meridian) also produces a series of analogous lines (isobars) characterizing the vertical pressure distribution in the given plane.

At the earth's surface (or rather at sea level) isobars are drawn according to data of observations at meteorological stations either for specific times (on synoptic charts) or from pressure means for various periods—month, season or year (on climatological charts). The pressure observed at meteorological stations, which are usually situated at various elevations, is reduced to sea level by means of the barometric formula. Isobars are plotted on the charts at definite intervals of pressure, usually 5 mb.

The spacing of isobars on these charts (surface or upper-air) provides an idea of the horizontal pressure variation. Quantitatively the magnitude of this variation is characterized by the horizontal pressure gradient $G_n = -\frac{\partial p}{\partial n}$ which indicates the pressure change along the normal n to the isobar and dips in the direction of falling pressure. The pressure gradient is expressed in millibars per degree of longitude (111 km).

A chart showing the sea-level pressure distribution for the European USSR is given as an example (Figure 17).

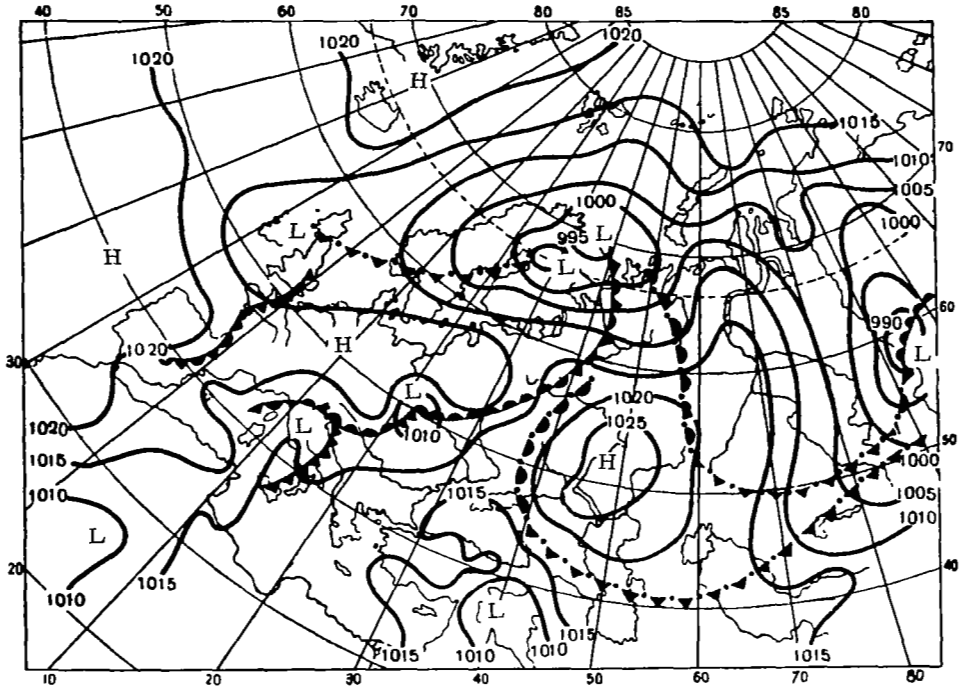


FIGURE 17. Surface weather chart. Isobars at sea level

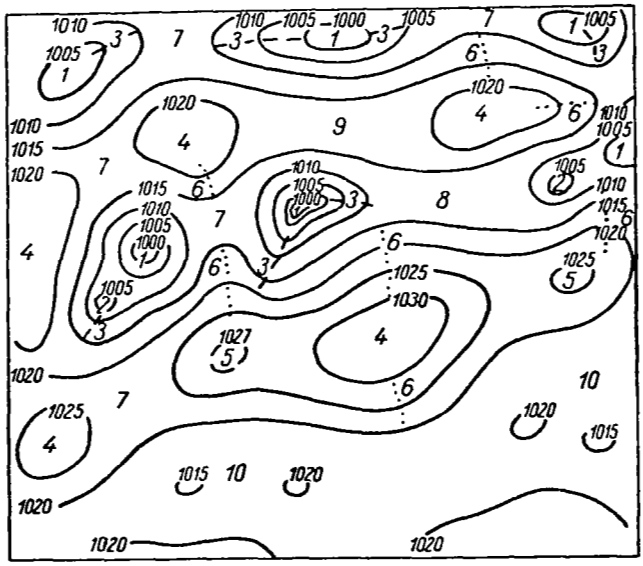


FIGURE 18. Forms of the pressure field

Despite the nonuniform distribution of pressure, it is always possible to single out regions of higher and lower pressure on isobaric charts of sufficiently large territories (Figure 17). Within the territory in question certain of these regions will be delineated by closed isobars and others by open ones. Depending on the wind system to which they are linked, closed pressure regions with low pressure at the center are called cyclones and those with high pressure at the center anticyclones. The former are denoted on charts by the letter L, the latter by H.

Obviously, the horizontal pressure gradient G is directed toward the center in cyclones and in the opposite direction in anticyclones. Further, it is usually greater in cyclones than in anticyclones.

Narrow regions of low pressure, elongate in some direction, are frequently found along the edges of cyclones. These elongate regions, usually delineated by open isobars, are called troughs. Elongate regions of high pressure, called ridges or wedges, are often found on the edges of anticyclones. The region between two pairs of cyclones and anticyclones is called a saddle. These basic forms of pressure relief are shown schematically in Figure 18.

It should be borne in mind that the isobars on a plane merely reflect the character of the pressure distribution in space. In the complex form isobaric surfaces have, as it were, their own relief and topography— they rise over high-pressure regions and dip over low-pressure regions.

§ 10. Mean distribution of pressure at the earth's surface

The mean distribution of pressure at the earth's surface (sea level) has been studied thoroughly on the basis of mean long-term data. In very general form, low pressure is observed in the equatorial zone. Poleward from this zone the pressure rises and reaches a peak at latitude 30–40° north and south. These are the so-called subtropical high-pressure belts. Further on toward the high latitudes (60–70°) the pressure drops again (especially in the southern hemisphere). And, finally, from this subpolar zone of low pressure to the poles the pressure again increases somewhat.

This, schematically, is the overall picture of pressure distribution over the earth. In reality even in average results it is considerably more complicated, as one can see, for instance, from isobaric charts (Figures 19 and 20) of the mean sea-level distribution in January (winter) and July (summer).

The uneven and complex distribution of pressure over the earth's surface is determined by factors of a thermal and also dynamical nature. The former include first of all the thermal influence of the earth's surface. It is manifested in the fact that conditions over cool surfaces tend to raise the pressure near the surface while, on the contrary, conditions over strongly heated surfaces tend to lower it. As a result high-pressure regions form over continents in the cold part of the year, then weaken or even vanish completely in summer, at which time they tend to form over oceans.

In addition to thermal causes, dynamical factors are also at work. As a result it is possible to have an inflow of air (i.e., pressure rise) in some

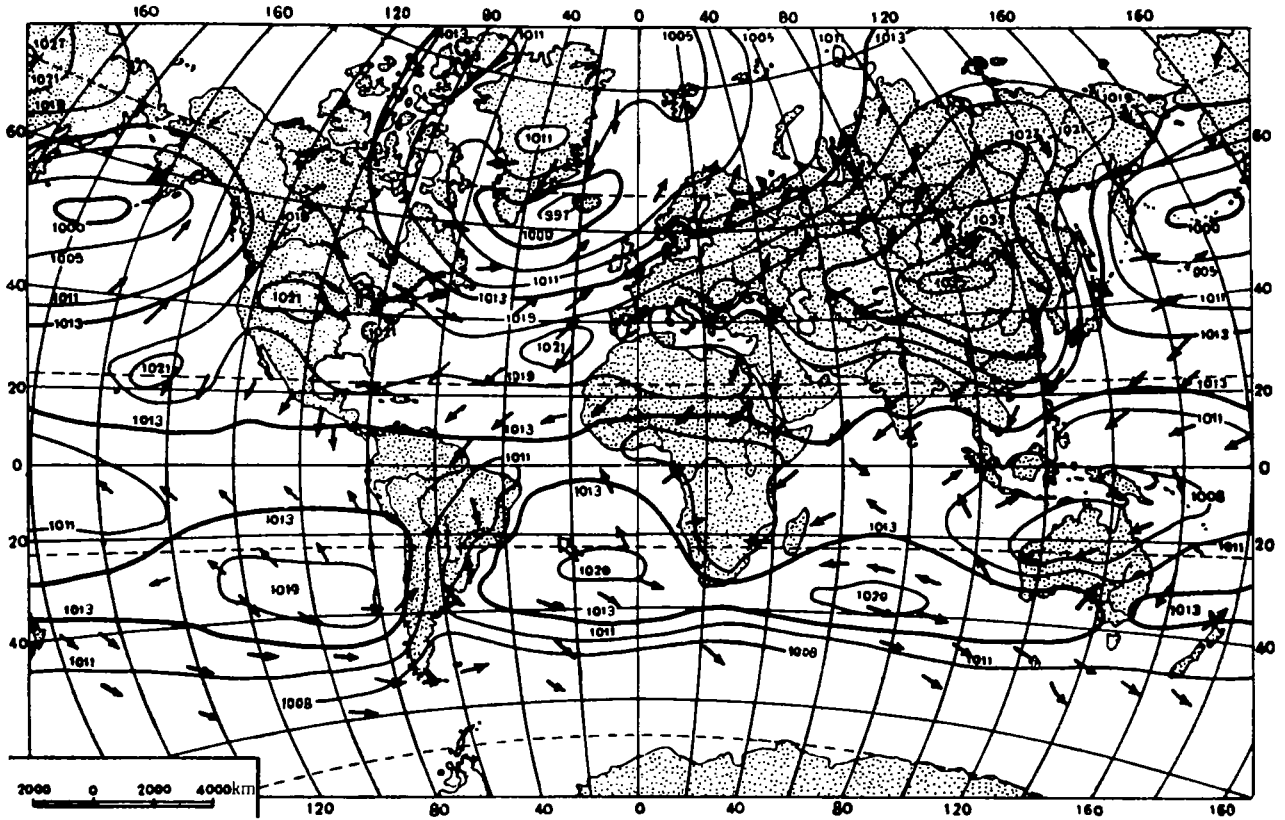


FIGURE 19. Isobars at sea level. January

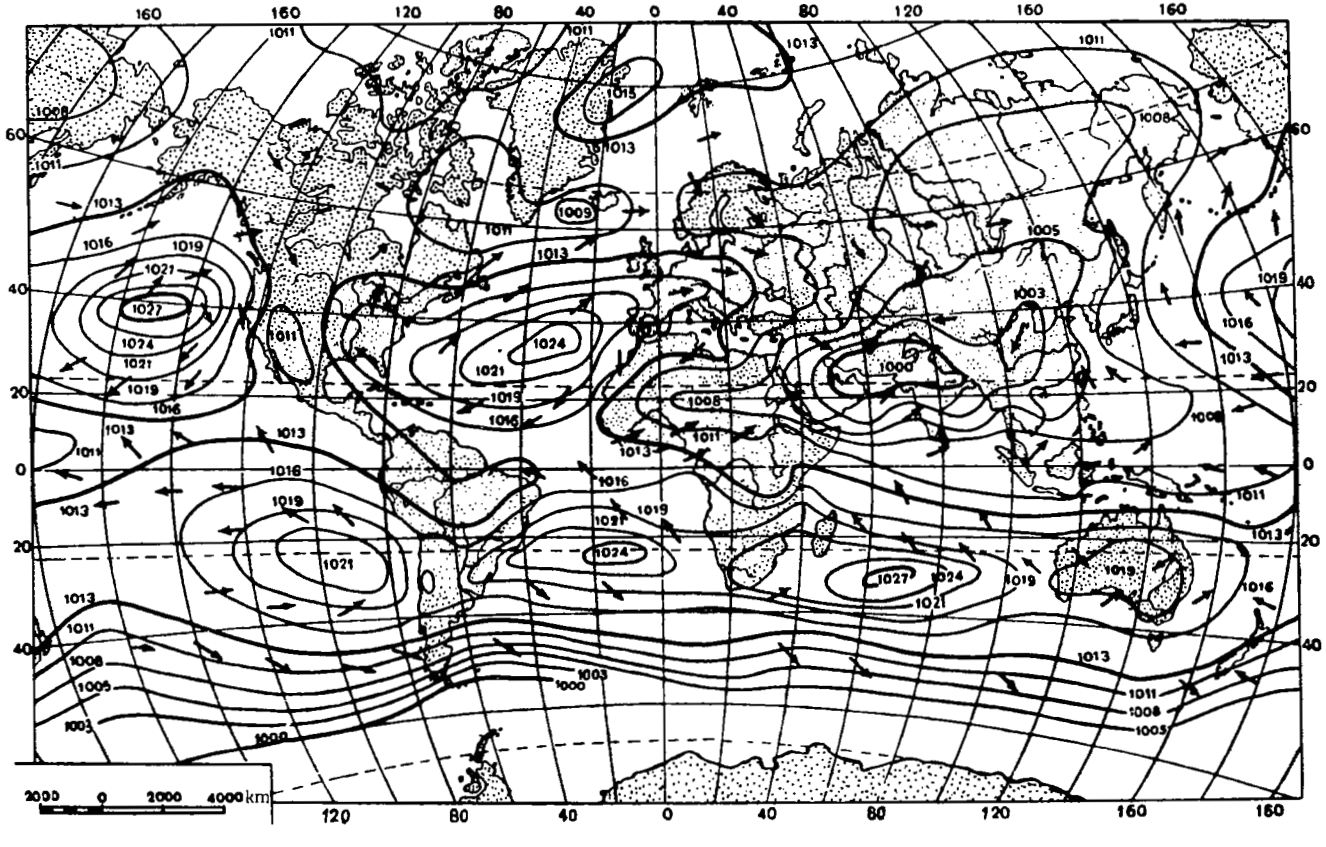


FIGURE 20. Isobars at sea level. July

regions and zones and, on the contrary, an outflow of air (i.e., pressure drop) in other regions.

The two factors act together and determine the actually observed distribution of pressure. Owing to the greater uniformity of structure of the earth's surface in the southern hemisphere the pressure distribution there is more regular than in the northern hemisphere and corresponds more closely to the scheme outlined above.

The high and low pressure belts indicated in the scheme break down into individual elements consisting of closed regions of high and low pressure. Some of these regions persist the year round, growing stronger or weaker in various seasons. Others are present in definite areas either in winter or in summer. These pressure regions have a great deal of influence on air currents, weather conditions and climate over a considerable territory and are therefore called centers of atmospheric activity.

Comparing the isobaric charts for July and January one might also notice that horizontal pressure gradients are greater in winter than in summer.

It should be stressed that at any particular time the differences in the distribution of pressure over the earth's surface are significantly larger than on the average. This is particularly true of the middle and high latitudes. The cyclones and anticyclones continually forming and decaying here remain neither stationary nor constant in form.

§ 11. Geopotential. The barometric formula for the geopotential

When studying air motion and the distribution of pressure in the atmosphere it is often convenient to introduce the gravity potential. The gravity potential is known in meteorology as the geopotential.

We recall that gravity potential is the term applied to a certain bounded continuous function of the coordinates of a point, the partial derivative of which in any direction will give the force acting at the given point in this direction. The potential characterizes the potential energy of the air particle situated at the given point.

To any point in the atmosphere (coordinates x, y, z) corresponds a unique value of the gravity potential $\Phi(x, y, z)$. Setting $\Phi(x, y, z) = \text{const} = A$, we obtain the equation of a certain surface at all points of which the potential has the same value. Such a surface is termed an equipotential, or level, surface. In meteorology it is called the geopotential surface.

By assigning various numerical values to the constant A in the above expression, we obtain a series of geopotential surfaces.

If from a certain point situated on the level surface $\Phi = A$ we move along the outward normal to a neighboring point, for which the value of the potential will be $\Phi + d\Phi = A_1$, then in order to transport a unit mass from the first surface to the second it is necessary to perform the work

$$d\Phi = g\,dz, \tag{41}$$

where g is the acceleration of gravity and dz the distance along the normal between the two level surfaces in question.

If one takes a certain constant value of $d\Phi$ it then follows from (41) that: 1) the distance dz between two neighboring level surfaces is inversely proportional to the active force g, i.e., when g is greater dz is smaller and vice versa; 2) dz cannot be zero, i.e., two level surfaces (different values of A) can neither intersect nor touch each other (or, equally, through every point in space there passes one and only one level surface). Since the acceleration of gravity decreases equatorward from the poles, it is clear from what we have said that the distance between two level surfaces increases away from the poles. Moreover, these surfaces do not coincide in the atmosphere with the surfaces of equal altitude (Figure 21).

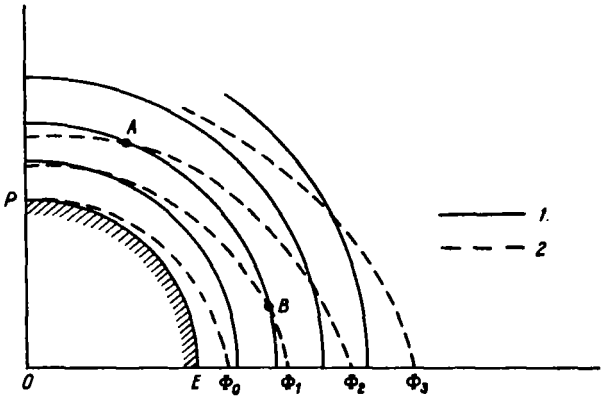


FIGURE 21. Location of equal-altitude (1) and equipotential (2) surfaces in the atmosphere

From Figure 21 it is evident that at different latitudes (points A and B) we have different values of the potential (Φ_1 and Φ_2) for the same height above the ground. Since every material point when left to itself tends to move toward the point with greater values of the gravity potential, a force tending to displace a material point from the pole to the equator is created along the surfaces of equal altitude.

If the value of the potential at the sea-level surface is arbitrarily assumed to be zero then the value of the potential at the level z can be found from equation (41).

Integrating the latter between the limits $(0, \Phi_z)$ and $(0, z)$, we have

$$\Phi_z = gz. \tag{42}$$

The unit of measurement for the geopotential is $1\ cm^2/sec^2$, but owing to its smallness a different unit, called the dynamical meter*, has been introduced in meteorology for practical purposes. This unit is 10^5 times larger, i.e., is $10^5\ cm^2/sec^2 = 10\ m^2/sec^2$. Its name derives from the fact that if one takes $g = 10\ m/sec^2$ in formula (42) and expresses z in meters, the value of the geopotential Φ in these units ($10^5\ cm^2/sec^2$) will be numerically equal to the height z expressed in geometric meters.

* This unit was used until 1950.

The relationship between the height of a certain point z and the value of the geopotential at this point allows us to characterize the position of the point by the value of the geopotential at this point (instead of the height expressed in meters). This is called the dynamical, or geopotential, height H.

Obviously, if we express H in dynamical meters we have

$$H \text{ dyn.m} = \frac{1}{10}\Phi = \frac{g}{10}z, \tag{43}$$

where z is in geometric meters.

Since g is smaller than $10\,\text{m/sec}^2$ and can be regarded as $g = 9.8\,\text{m/sec}^2$ on the average, there is a difference between the height of the point expressed in dynamical meters, H dyn m, and the same height expressed in geometric meters, z m. Although this difference is small (of the order of 2 % on the average) for large values of the height it may prove significant in absolute magnitude (of the order of tens or even hundreds of meters). Accordingly a new unit has recently (since 1950) been introduced for the geopotential in place of the dynamical meter. This is the so-called g e o p o t e n t i a l m e t e r (gp m), numerically very close to the geometric meter.

The geopotential meter is defined as the quantity of work which must be done to move a unit mass over a distance of one meter against the force of gravity for a value of the acceleration of gravity of $g = 9.8\,\text{m/sec}^2$, i. e., $1\,\text{gp m} = 9.8\,\text{m}^2/\text{sec}^2$.

Obviously, 1 gp m = 0.98 dyn m and 1 dyn m = 1.02 gp m. Consequently, at a certain height z m the value of the geopotential expressed in geopotential meters is $H\,\text{gp m} = \frac{gz}{9.8}$, and since the acceleration g at various points of the earth's surface and atmosphere differs from the average value $g = 9.8\,\text{m/sec}^2$ by no more than 0.5 %, it is evident that H gp m is numerically very close to z m (in practice H gp m and z m can be regarded as equal).

The introduction of the geopotential makes it possible to relate the variation in pressure with height to the variation of the geopotential. To do this, let us turn to the equation of hydrostatics $dp = -\rho g dz$ and, bearing (42) in mind, write it in the form

$$dp = -\rho d\Phi. \tag{44}$$

Since $\rho = \frac{p}{RT_v}$,

$$d\Phi = -RT_v\frac{dp}{p}, \tag{45}$$

whence

$$\int_{\Phi_1}^{\Phi_2} d\Phi = -RT_v\int_{p_1}^{p_2}\frac{dp}{p}, \tag{46}$$

where Φ_1 and Φ_2 are the values of the geopotential at the levels z_1 and z_2, at which the pressure is respectively p_1 and p_2.

Introducing the mean virtual temperature $\overline{T}_v$, we find from (46) that

$$\Phi_2 - \Phi_1 = R\overline{T}_v \ln\frac{p_1}{p_2}. \tag{47}$$

Taking $\Phi_1 = 0$ and $p_1 = p_0$ at sea level, we write the following for the geopotential $\Phi_2 = \Phi_z$ at a certain height, at which the pressure $p_2 = p$:

$$\Phi_z = R\overline{T}_v \ln\frac{p_0}{p}. \tag{48}$$

The relation (47) (and also (45)) is entirely analogous to the barometric formula introduced earlier (19) and may be called the barometric formula for the geopotential. It relates the difference between the values of the geopotential at two points to the values of the pressure at these points and to the mean virtual temperature of the layer of air between them.

This formula is highly convenient for comparing atmospheric pressure at various points. We mentioned earlier that the value of the geopotential at a certain point is known by convention as the dynamical, or geopotential, height and is denoted by H. The barometric formula for the geopotential (47) may therefore be written in a convenient form for practical calculations:

$$H_2 - H_1 = 0.102 \cdot 2.3 \cdot 287 \overline{T}_v \log \frac{p_1}{p_2} = 67.44 \overline{T}_v \lg \frac{p_1}{p_2} \text{ gp m.} \qquad (49)$$

Here we have replaced the natural logarithm by common logarithms and have set $R = 287 \text{ m}^2/\text{sec}^2 \cdot \text{deg}$. The factor 0.102 results from the transition to geopotential meters.

One can transform formula (48) analogously:

$$H_p = 67.44 \overline{T}_v \log \frac{p_0}{p} \text{ gp m.} \qquad (50)$$

§ 12. Absolute and relative geopotential. Charts of pressure topography

Consider a certain definite isobaric surface — i. e., at all points of which the pressure has the same value p — in the atmosphere. The elevation of this surface above sea level will be different at different points. The value of the geopotential at each point of this surface, or its geopotential height H_p, is called the absolute geopotential of the given isobaric surface p and is given by the relation (48)

$$\Phi_p = R\overline{T}_v \ln \frac{p_0}{p} = H_p.$$

One can see from this relation that H_p for the given isobaric surface ($p = \text{const}$) depends on the mean virtual temperature $\overline{T}_v$ of the air layer, the height of the isobaric surface above sea level and the value of the surface pressure p_0. The dependence on each of these quantities is direct. We can find the variation of the absolute geopotential H_p for varying $\overline{T}_v$ and p_0 from (48) by taking the logarithmic derivative

$$\frac{d\Phi_p}{\Phi_p} = \frac{d\overline{T}_v}{\overline{T}_v} + \frac{\frac{dp_0}{p_0}}{\ln \frac{p_0}{p}}$$

or, since $\Phi_p = R\overline{T}_v \ln \frac{p_0}{p}$,

$$d\Phi_p = R \ln \frac{p_0}{p} d\overline{T}_v + R\overline{T}_v \frac{dp_0}{p_0}. \qquad (51)$$

From this formula it follows that when $\overline{T}_v$ rises or drops (for $p_0 = \text{const}$) the absolute geopotential (height of the isobaric surface) also increases or decreases. Further, this increase is greater for isobaric surfaces

situated at higher levels (smaller p). When $\overline{T}_v$ increases by 1° the isobaric surface at, say, $p = 700$ mb (for $p_0 = 1000$ mb) is raised by $80\,m^2/sec^2$. When the surface pressure p_0 increases by 1 mb all equipotential surfaces are raised by approximately $80\,m^2/sec^2$.

Aside from the absolute geopotential (geopotential height) one may also consider the so-called relative geopotential, by which is meant the difference between the geopotentials (heights) of two isobaric surfaces p_1 and p_2, i.e., the difference $H_{p_1} - H_{p_2}$. It is customarily denoted by $H_{p_1}^{(p_2)} = H_{p_1} - H_{p_2}$. As one can see from formulas (47) and (49), the relative geopotential depends only on the mean virtual temperature of the layer between the isobaric surfaces. It increases with increasing $\overline{T}_v$ and decreases with decreasing $\overline{T}_v$. Thus for $p_1 = 1000$ mb and $p_2 = 500$ mb the value of the relative geopotential can be calculated from the formula obtained from (50)

$$H_{500} - H_{1000} = H_{1000}^{500} = 67.44\overline{T}_v \log\frac{1000}{500} = 20.3\overline{T}_v \text{ gp m.} \qquad (52)$$

i.e., a 1° change in $\overline{T}_v$ for this layer will produce a change of about 20 gp m in its thickness.

The concepts of the absolute and relative geopotential are widely used in constructing so-called charts of pressure topography.

Let us take any isobaric surface (p = const) lying over a certain more or less extensive territory. We connect points with the same value of the absolute geopotential (geopotential height of the given surface above sea level) by means of lines. These lines are called altitude isograms. From what we said earlier it is clear that wherever the elevation above sea level of an isobaric surface is greater its geopotential height will also be greater, and vice versa, i.e., a chart of altitude isograms is a representation of the relief of the isobaric surface. Such charts are called charts of absolute topography. They are denoted by the symbol AT with a subscript indicating the isobaric surface to which they refer. Thus AT_{500} designates the chart of absolute topography of the 500 mb surface.

Usually AT charts are compiled for so-called principal isobaric surfaces, chosen by convention at pressures of 1000, 850, 700, 500, 300, 200 and 100 mb. Thus AT charts provide a graphic characterization of the distribution of pressure in the atmosphere. Examples of absolute-topography charts are given in Figures 22 and 23 for the 500 mb surface.

Aside from AT charts, there are also charts of relative topography, denoted by $RT_{p_1}^{p_2}$, which show the difference in geopotential height between two selected isobaric surfaces (lower one with the pressure p_1 and upper one with the pressure p_2). RT charts are illustrated in Figures 24 and 25. The isograms on $RT_{p_1}^{p_2}$ charts indicate the excess of the upper isobaric surface p_2 over the lower surface p_1. As one can see from (50), the isograms also characterize the values of the mean virtual temperature of the layer between the surfaces in question. Further, the height difference expressed in decameters is almost exactly twice the value of the virtual temperature.

In practice RT_{1000}^{500} charts are the ones usually constructed. These charts characterize the atmosphere practically from the sea-level surface (p = 1000 mb) to an altitude of the order of 5–5.5 km (p = 500 mb), i.e., a layer containing nearly half the entire mass of the atmosphere.

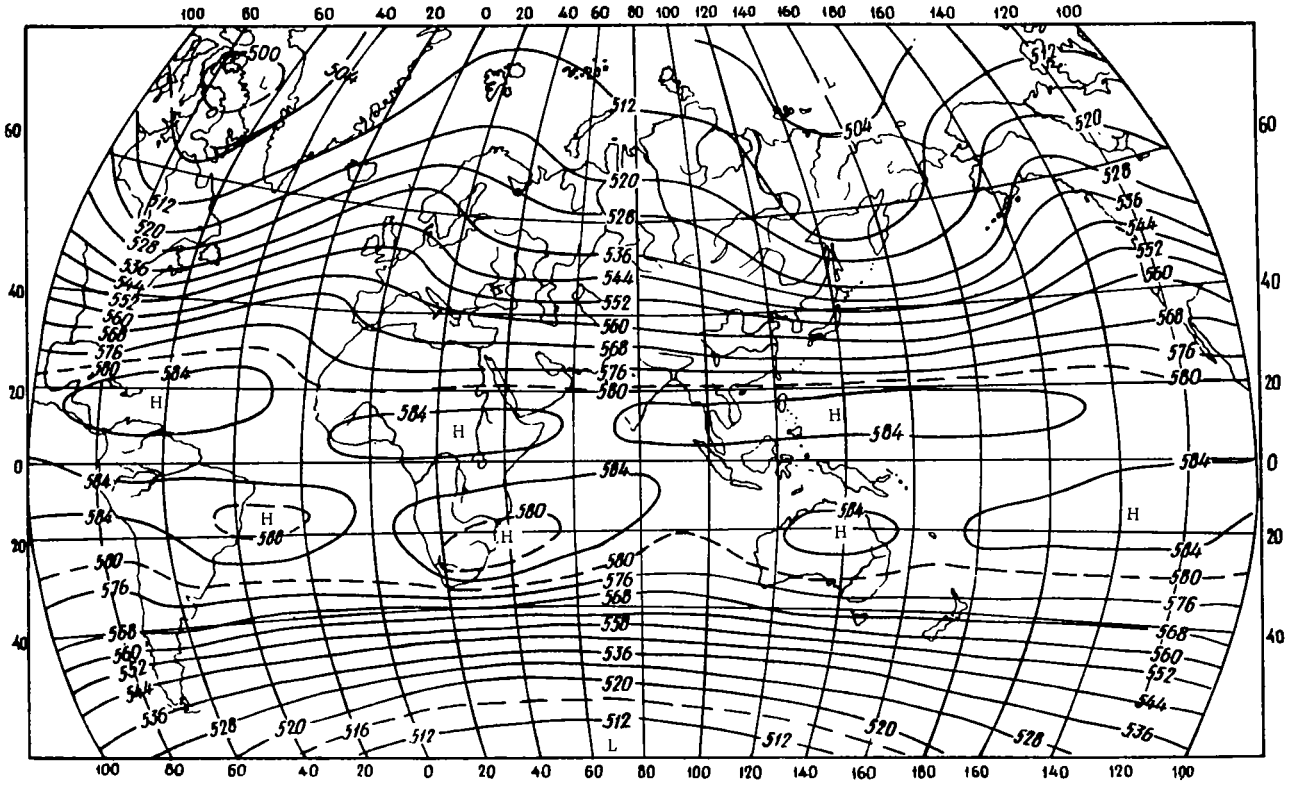


FIGURE 22. Absolute topography of 500 mb surface (AT_{500}). December-February (after Kh. P. Pogosyan).

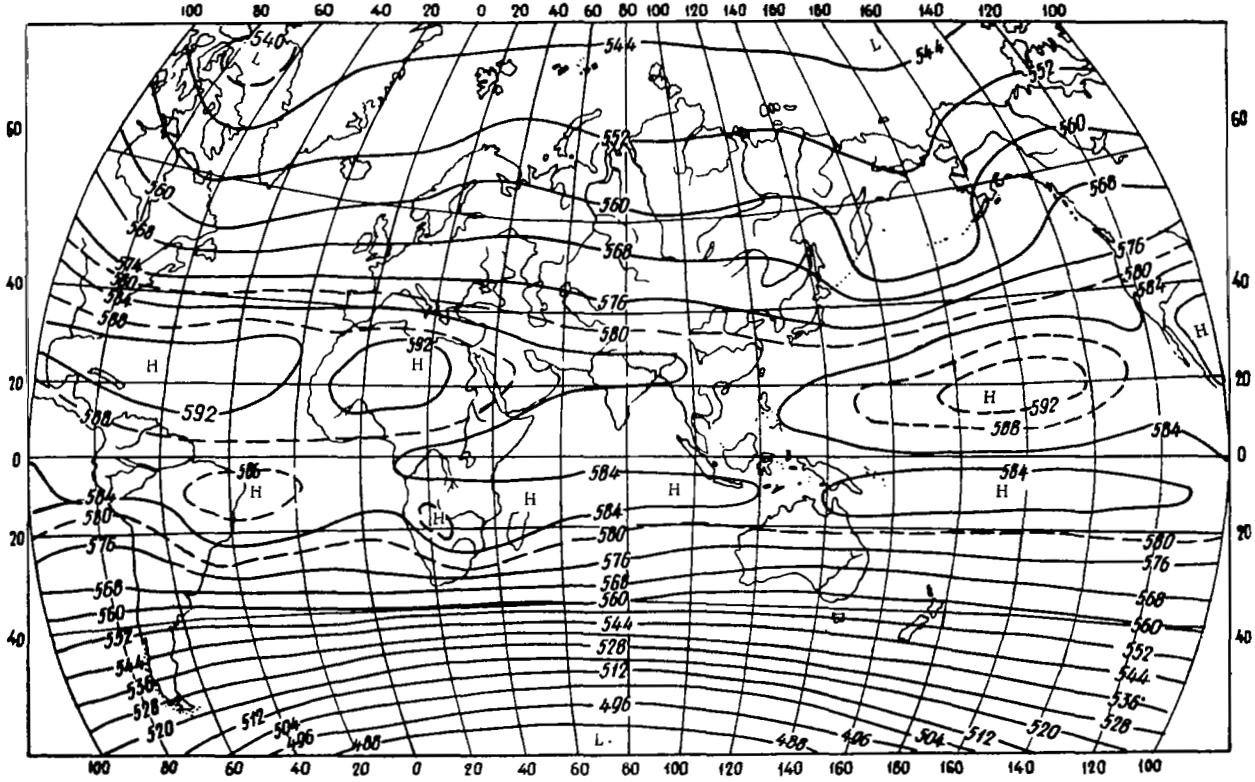


FIGURE 23. Absolute topography of 500 mb surface (AT_{500}). July-August (after Kh. P. Pogosyan).

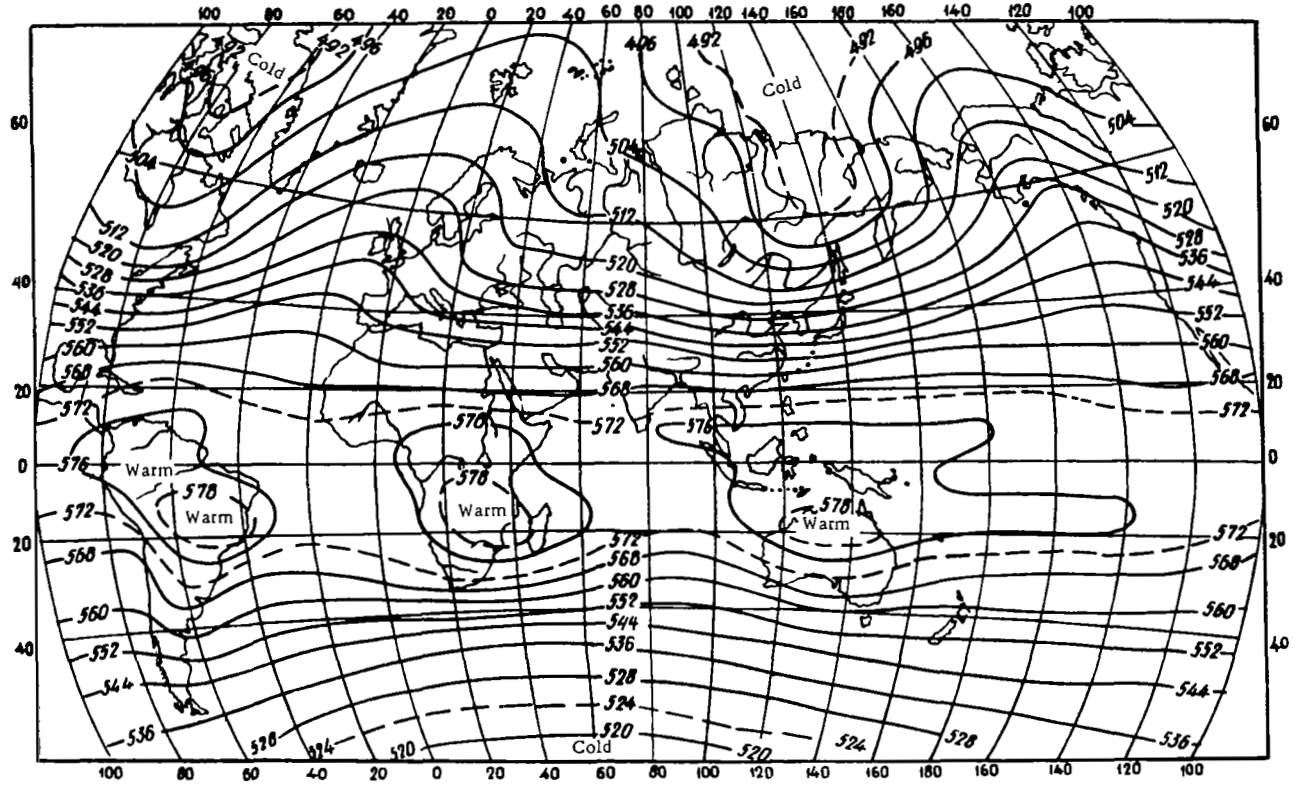


FIGURE 24. Chart of relative topography (OT^{500}_{1000}). December-February (after Kh.P. Pogosyan)

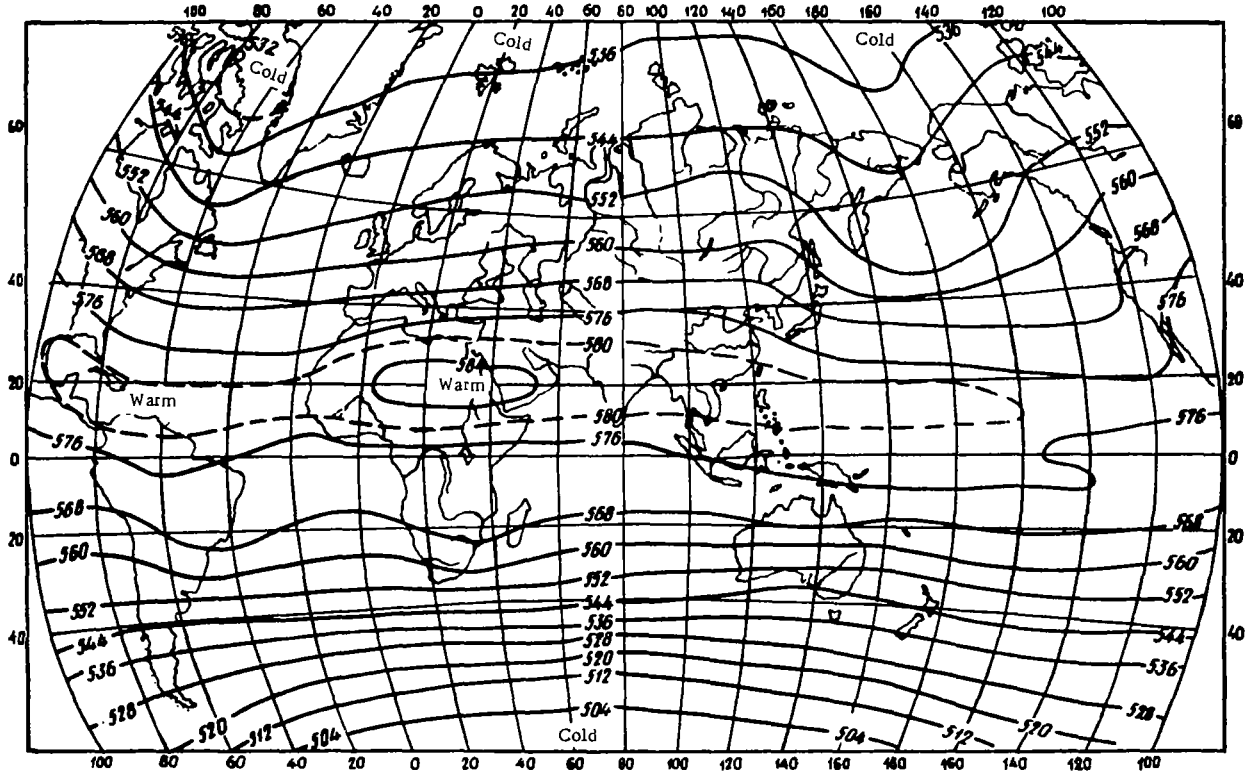


FIGURE 25. Chart of relative topography (OT_{1000}^{500}). July-August (after Kh.P. Pogosyan)

AT and RT charts are of great value for studying the development of atmospheric processes and are used particularly extensively in synoptic meteorology.

§ 13. Pressure distribution in the atmosphere at different levels

As a result of the vertical variation of pressure, which is closely related to the vertical distribution of air temperature, the pressure distribution at any level above the earth's surface differs sharply from the distribution of surface pressure. In characterizing this distribution it is very convenient to employ charts of absolute topography.

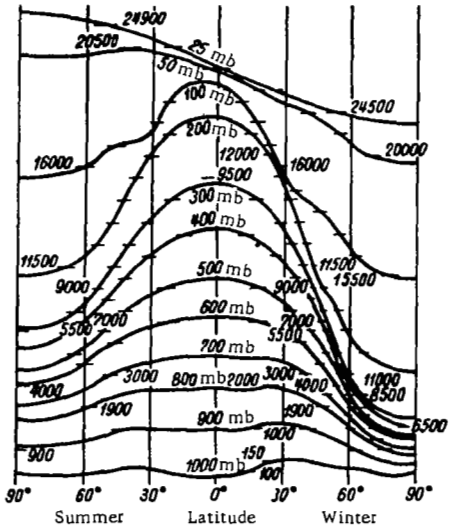


FIGURE 26. Geopotentials of the principal isobaric surfaces (heights are in dynamical meters)

The mean altitudes of the principal isobaric surfaces (in dynamical meters) for summer and winter are given in Figure 26; AT_{500} charts for the entire northern hemisphere for the same seasons are given in Figures 22 and 23. Analysis of these figures shows that features of the surface pressure distribution stemming from the nonuniformity of the ground gradually fade with increasing altitude. Furthermore, the surface pattern of high- and low-pressure belts described earlier disappears at the height of the 700 mb isobar. Above this level and up to the 100 mb surface, the region of highest pressure lies over the equator and pressure drops away from the equator toward the poles; a low-pressure region occurs at the poles. The spacing of contours on AT charts and the inclination of isobaric surfaces are indications of the magnitude of the horizontal pressure gradient. It can be seen that the gradients on the charts given here are considerably steeper in winter than in summer.

We note once again that mean charts always attenuate and obscure many details and that the actual pressure distribution at different heights in the atmosphere at any single instant can be substantially different from the one shown.

It should be emphasized that air movements in the atmosphere are very closely linked to the distribution of pressure.

§ 14. Variation of pressure in time

Atmospheric pressure does not remain constant in time at every given point. The variations of pressure are mainly nonperiodic and are caused by processes of general circulation of the atmosphere, which are responsible for a continuous flow of pressure formations of various kinds (cyclones, anticyclones, etc.) past each given point. The range of variation of pressure is fairly broad. Thus observations have established that variations in middle and high latitudes are mainly from 970 to 1040 mb, reaching 925 mb (minimum) and 1070 mb (maximum) in exceptional cases.

Periodic fluctuations, annual as well as daily, can also be established against the background of these nonperiodic changes by statistical processing of the observational materials. These are governed by stable factors.

Annual fluctuations of pressure depend largely on physico-geographical conditions: they are small in the equatorial zone and larger toward the high latitudes. The annual march of pressure is different over oceans and continents. Under continental conditions annual pressure amplitudes are significantly larger than over the oceans, and the pattern itself is reversed: over continents the pressure maximum is observed in winter and the pressure minimum in summer, while the opposite occurs over the oceans. The annual amplitude of pressure fluctuation is about 10 mb. Thus it is 8–12 mb over the European USSR. Particularly large amplitudes are observed in the central regions of continents at large distances from the coast. In Siberia, for example, the annual pressure amplitude reaches 25–30 mb. The annual amplitude decreases in the direction of the ocean shore and is smaller over middle-latitude oceans, amounting to a few millibars. Annual pressure fluctuations are also observed in the free atmosphere. Their amplitude there increases markedly with altitude while the pattern reverses itself starting from a limited height (pressure peak in summer and minimum in winter over continents).

The daily march of pressure is characterized by very limited amplitudes. These are especially small in middle and high latitudes, where they amount to a few tenths of a millibar. In the tropics, where the amplitude of daily fluctuations is somewhat larger (reaching 3-4 mb), daily fluctuations can even be detected directly from 24 hr records of pressure variations. In higher latitudes, where the nonperiodic fluctuations are large, daily fluctuations are not so easily detected and special, complicated processing of observational data over long periods is necessary. The daily pressure march displays two maxima (9–10 and 21–22 hrs local time) and two minima (3–4 and 15–16 hrs).

To explain the daily march of pressure the curve obtained from observations is subjected to processing by the method of harmonic analysis. Waves with periods of 24, 12, 8 and 6 hrs are singled out. It has been

established that the daily wave (24 hr period) manifests itself conspicuously and regularly in equatorial and tropical regions. In high latitudes it is less regular and depends on the character of relief. The half-day wave (12 hr period) is more clear-cut throughout.

These small daily fluctuations in pressure have no practical significance but their study is of a very considerable interest from the theoretical point of view. There exists as yet no accomplished theory but it is fairly well established that they are due to elastic fluctuations of the atmosphere. They may be a consequence of the periodic heating of the atmosphere by the sun's rays and associated expansion of the air on the side-illuminated by the rays. Furthermore, as a result of the influence of the tidal force exerted by the sun and moon, the atmosphere is the site of ebb-flow movements analogous to those observed in the sea. This is felt in the half-day fluctuations of pressure. Indeed, it has been established that the half-day wave contains components with both solar (12 hrs) and lunar (12 hrs 25 min) periods, i.e., solar and lunar half-day waves. The latter, however, are very faint and their amplitude is approximately 1/16th of the amplitude of the solar half-day fluctuations; over Western Europe, for example, they constitute in all about 0.01–0.04 mb. These fluctuations are so small that they cannot perceptibly affect processes taking place in the troposphere.

Chapter 6

FUNDAMENTALS OF ATMOSPHERIC THERMODYNAMICS

When we study the processes taking place in the atmosphere we are obliged to deal with the transformation of one form of energy into another. Such a transformation takes place in the heating and cooling of air, absorption of radiant energy, condensation of water vapor, and so on. In these processes the internal parameters characterizing the state of the system which we are considering (unit mass of air) undergo changes. Important interferences concerning these changes follow from the fundamental tenets of thermodynamics.

In the present chapter a number of thermodynamic problems is considered with reference to atmospheric processes.

§ 1. Fundamental equations

We recall that the first law of thermodynamics is the general expression of the law of conservation of energy. Out of the great number of formulations of this law we will employ the one which states that the quantity of heat ΔQ imparted to a system is used in: 1) increasing its internal energy (ΔU_1) and 2) in doing work against external forces (ΔU_2). This can be written as

$$\Delta Q = \Delta U_1 + \Delta U_2 . \tag{1}$$

Let us consider a unit mass (1 g) of dry air, regarding it as an ideal gas. Further, let us assume that during the process under consideration the (external) pressure $\overline{p}$ exerted by the surrounding medium on the system in question (mass of gas) is equal to its internal pressure p, which is determined by molecular motion in the mass itself. Such processes are called quasi-static. Experience shows that in the atmosphere the pressure of a gas in a small volume rapidly becomes equal to the pressure of the surrounding medium and thus the above condition is fairly well fulfilled. It is not fulfilled only in the upper atmospheric layers where pressure pulsations are very considerable, and also for very high velocities approaching the velocity of sound.

Under these conditions the change in the internal energy of one gram of air ΔU_1 can be written in the form

$$\Delta U_1 = c_v \Delta T, \tag{2}$$

where c_v is the specific heat at constant volume.

If the only external force we consider is the pressure, then the work

done by the system at constant pressure p is given by

$$\Delta U_2 = p\Delta v,$$

or, in thermal units,

$$\Delta U_2 = Ap\Delta v, \tag{3}$$

where $A = 0.24 \cdot 10^{-7}$ cal/erg is the thermal equivalent of work and v is the specific volume.

Inserting (2) and (3) into (1) and passing to differentials, we write the latter as

$$dQ = c_v dT + Apdv. \tag{4}$$

This equation can be written in a different form which is often useful. To do this we differentiate the equation of state

$$pv = RT,$$

whence we find

$$pdv = RdT - vdp.$$

Introducing (5) into (4), we obtain

$$dQ = (c_v + AR)\,dT - Avdp.$$

Since $c_p - c_v = AR$,*

$$dQ = c_p dT - Avdp. \tag{4'}$$

Introducing $v = \frac{RT}{p}$ into (4'), we obtain

$$dQ = c_p dT - ART\frac{dp}{p}. \tag{4''}$$

The relations (4), (4') and (4'') are different expressions of the first law of thermodynamics. We recall that c_p and c_v are practically independent of the temperature and that $c_p = 0.2388 \approx 0.24$ cal/g·deg, $c_v = 0.1712 \approx$ ≈ 0.17 cal/g·deg and $\frac{c_p}{c_v} = x = 1.41$.

Let us note especially that, according to (4),

$$dT = \frac{dQ}{c_p} + \frac{Av}{c_p}dp, \tag{5}$$

i.e., the temperature of a system can be changed not only by the supply (or removal) of a quantity of heat dQ but also by a change in the external pressure dp. If the pressure increases $(dp>0)$ then for the same dQ the temperature of the system will be higher; conversely, if the pressure decreases $(dp<0)$ the temperature will be lower. In the particular case where $dQ=0$, the temperature change is completely determined by the variation of pressure

$$dT = -\frac{Av}{c_p}dp. \tag{6}$$

* This follows from the fact that for p = const $dp=0$ and $dQ = c_p dT$; consequently, $c_p dT = (c_v + AR)dT$, whence we arrive at Mayer's equation $c_p - c_v = AR$.

§ 2. Polytropic changes of state. The equation of polytropy

The transition of a given mass of gas from one state to another can take place along various paths. The fundamental parameters which characterize the state of the gas (p, $v=\frac{1}{\rho}$ and T) can change in various ways and every path of transition can be characterized with the help of an equation relating these quantities. The quantity of heat imparted to the system can also vary.

In meteorology the two most important processes are: the adiabatic process, in which the influx of heat dQ is zero, and the polytropic process, in which the influx of heat dQ is proportional to the temperature increment dT, i.e.,

$$dQ=cdT, \tag{7}$$

where c is a proportionality factor which carries the meaning of specific heat and is called the polytropic heat capacity.

We will first consider the more general polytropic process. Setting the right-hand sides of (7) and (4'') equal, we write

$$(c_p-c)\frac{dT}{T}=AR\frac{dp}{p} \quad \text{or} \quad \frac{dT}{T}=\frac{AR}{c_p-c}\frac{dp}{p}. \tag{8}$$

Integrating (8) from the initial state (p_0, T_0, ρ_0) to the final state (p_1, T_1, ρ_1) we obtain

$$\frac{T_1}{T_0}=\left(\frac{p_1}{p_0}\right)^{\frac{AR}{c_p-c}}. \tag{9}$$

Since $T=\frac{pv}{R}$, and $\frac{T_1}{T_0}=\frac{p_1v_1}{p_0v_0}$, we have, instead of (9),

$$\frac{p_1v_1}{p_0v_0}=\frac{p_1}{p_0}^{\frac{AR}{c_p-c}}$$

or

$$\frac{v_1}{v_0}=\left(\frac{p_1}{p_0}\right)^{\frac{AR}{c_p-c}-1}. \tag{10}$$

Introducing $AR=c_p-c_v$, we obtain

$$\frac{v_1}{v_0}=\left(\frac{p_1}{p_0}\right)^{\frac{c-c_v}{c_p-c}}=\left(\frac{p_0}{p_1}\right)^{\frac{1}{k}}, \tag{11}$$

with $k=\frac{c_p-c}{c_v-c}$, and thus

$$p_0v_0^k=p_1v_1^k=\ldots=pv^k=\text{const}. \tag{12}$$

This is the well-known equation of polytropy, where k is the so-called polytropic exponent.

By assigning values from $+\infty$ to $-\infty$ to the parameter c, the equation of polytropy makes it possible to describe a broad class of different (polytropic) processes. The following very simple particular cases are of interest:

1) $c=c_p$; then $p=\text{const}$, i.e., on transition from one state to another the pressure remains constant. This is the so-called isobaric process. It is described by the condition $k=0\,(dQ=c_pdT)$;

2) $c=\infty$; then $T=\text{const}$ and we have an isothermal process for $k=1\ (dT=0)$;

3) $c=c_v$; then $v=\text{const}$ and we have an isosteric process for $k=\infty\ (dQ=c_v dT)$;

4) of especially great importance is the adiabatic process mentioned above; here the influx of heat is zero $(dQ=0)$, and $c=0$, while $k=\varkappa=\frac{c_p}{c_v}=1.4$.

We will examine this process in greater detail in the following section.

§ 3. Adiabatic processes. The Poisson equation

In many cases atmospheric processes can be treated approximately as adiabatic, i.e., occurring without any exchange of heat between the mass of air (particle) being considered and the surrounding medium. This is permissible whenever the change in the heat content $c_p dT$ due to the work of expansion is much greater than the influx of heat from the outside, and the latter can be disregarded. In particular, this holds in vertical movements of air when the pressure change is very considerable and the heat influx is small.

For adiabatic processes the equation of energy influx (4'') takes the form

$$dQ=c_p dT-ART\frac{dp}{p}=0, \tag{13}$$

or

$$\frac{dT}{T}=\frac{AR}{c_p}\frac{dp}{p}. \tag{13'}$$

Integrating from T_0 and p_0 to T_1 and p_1 we have

$$\frac{T_1}{T_0}=\left(\frac{p_1}{p_0}\right)^{\frac{AR}{c_p}}. \tag{14}$$

Bearing in mind that $AR=c_p-c_v$ and, consequently,

$$\frac{AR}{c_p}=\frac{c_p-c_v}{c_p}=\frac{\varkappa-1}{\varkappa},$$

where $\varkappa=\frac{c_p}{c_v}$ and $\frac{\varkappa-1}{\varkappa}=0.288$, we obtain instead of (14)

$$\frac{T_1}{T_0}=\left(\frac{p_1}{p_0}\right)^{\frac{\varkappa-1}{\varkappa}}=\left(\frac{p_1}{p_0}\right)^{0.288}. \tag{15}$$

This is the Poisson equation, or equation of the dry adiabatic. Let us write it as follows:

$$\left(\frac{T_1}{T_0}\right)^{\varkappa}=\left(\frac{p_1}{p_0}\right)^{\varkappa-1},$$

or

$$\frac{T_1^{\varkappa}}{p_1^{\varkappa-1}}=\frac{T_0^{\varkappa}}{p_0^{\varkappa-1}},$$

whence, bearing in mind that $p=\frac{RT}{v}$, we obtain

$$pv^{\varkappa}=\text{const}. \tag{16}$$

Obviously, this equation can also be obtained from the general equation of polytropy (12) by setting $k = \frac{c_p - c}{c_v - c} = \varkappa$ (which is what happens for $c = 0$).

In meteorology Poisson's equation is generally used in the form (15) for calculations.

§ 4. Dry adiabatic temperature gradient

Let us take equation (9), which relates the temperature to the pressure, and see how the temperature of a certain mass of air changes when we change the external pressure p; this question is particularly important when the mass in question (particle) is moving along the vertical. If this movement is sufficiently slow the process can be regarded as quasi-static, i.e., one can set $p^* = p$. We will distinguish the pressure and other parameters referring to the air particle by a star.

Taking the logarithm and differentiating formula (9), and also bearing in mind that $p_0 = \text{const}$ and $T_0 = \text{const}$, we write

$$\frac{1}{T^*}\frac{dT^*}{dz} = \frac{AR}{c_p - c}\frac{1}{p}\frac{dp}{dz}. \tag{17}$$

We make use of the hydrostatic equation (for the surrounding medium)

$$\frac{dp}{dz} = -\rho g = -\frac{pg}{RT} \tag{18}$$

and rewrite (17) (remembering that $p^* = p$) in the form

$$\frac{dT^*}{dz} = -\frac{Ag}{c_p - c}\frac{T^*}{T}. \tag{19}$$

This gives the change in temperature of a mass of air moving along the vertical.

Since the difference between the temperature of the air mass in question T^* and that of the surrounding medium T is usually comparatively small, one can set $\frac{T^*}{T} \approx 1$. We obtain, with a sufficient degree of accuracy,

$$\frac{dT^*}{dz} = -\frac{Ag}{c_p - c} = \frac{Ag}{c_p}\frac{1}{1 - \frac{c}{c_p}}. \tag{20}$$

In the case of an adiabatic process (where the polytropic heat capacity $c = 0$) we obtain instead of (19)

$$\frac{dT^*}{dz} = -\frac{Ag}{c_p}\frac{T^*}{T}, \tag{21}$$

and instead of (20)

$$\frac{dT^*}{dz} = -\frac{Ag}{c_p} = -\gamma_a. \tag{21'}$$

The quantity γ_a is known as the dry adiabatic temperature gradient.* It is a constant whose numerical value for $g = 980.6\ \text{cm/sec}^2$

* [Usually known as the "adiabatic lapse-rate" in Western literature; it is important to remember that whenever the author uses here the term "temperature gradient" he means the lapse rate, i.e., the negative of the actual temperature gradient. Other gradients, e.g., pressure, have the usual meaning.]

is given by

$$\gamma_a = \frac{Ag}{c_p} = \frac{0.239 \cdot 10^{-7} \cdot 980.6}{0.24} = 0.978 \cdot 10^{-4}\ ^\circ/\text{cm} \approx$$

$$\approx 0.98^\circ/100\ \text{m} \approx 1^\circ/100\ \text{m}. \qquad (21'')$$

Consequently, a rising mass of air containing unsaturated water vapor (for $g = 980.6\ \text{cm/sec}^2$) is cooled or warmed by 0.98° (or roughly 1°) for every 100 m of ascent or descent respectively.

From (20) it is obvious that the smaller the polytropic heat capacity c (the smaller the ratio $\frac{c}{c_p}$) the closer the process is to being adiabatic.

Making use of the dry adiabatic gradient one can write the following expression for an ascending mass of dry air

$$T_z^* = T_0^* - \gamma_a \frac{z}{100}, \qquad (22)$$

where T_z^* and T_0^* are its temperatures at the level z and at the initial level $z=0$, respectively; the height z is expressed in meters.

If we represent this result graphically in the system of coordinates z and T we will obtain the function $T^*(z)$ in the form of a straight line (forming an angle of 45° with the coordinate axes if the scale is such that 1° corresponds to 100 m). This line is called the dry adiabatic, or curve of state* of an ascending air particle.

§ 5. Potential temperature and its variation with height

Poisson's formula (15) is of considerable significance in meteorology as it relates the two fundamental parameters of state p and T. As an air mass moves along the vertical it enters layers with different pressures and its temperature will change independently of all other causes.

To compare the thermodynamic state of two air masses it is necessary to consider both p and T, which is inconvenient. Thus it is desirable to introduce a single characteristic. The characteristic used for comparing the state of air masses is called the potential temperature Θ. By this is meant the temperature which a mass of dry air will attain if it is brought adiabatically to the standard pressure (taken to be 1000 mb).

Making use of Poisson's equation and assuming, from the definition of potential temperature, that $T_0 = \Theta$ and $p_0 = 1000$ mb, we write

$$\frac{T^*}{\Theta} = \left(\frac{p}{1000}\right)^{\frac{AR}{c_p}}, \qquad (23)$$

whence

$$\Theta = T^* \left(\frac{1000}{p}\right)^{\frac{AR}{c_p}} = T^* \left(\frac{1000}{p}\right)^{0.288}. \qquad (24)$$

With the above one can calculate Θ for any value of p and T. Practical calculations are generally carried out with the help of prepared tables.

Approximate estimates can be made with the help of the dry adiabatic gradient using the relation (22). Indeed, it follows from (22) that when an

* [See also p. 118.]

air mass descends adiabatically from the height z to $z=0$ its temperature T_z^* will rise by $\gamma_a \frac{z}{100} \approx \frac{z}{100}$ $(\gamma_a = 1)$. But usually at $z=0$ the pressure is close to 1000 mb and therefore when a particle descends to this level its temperature T^* will be approximately equal to Θ, i.e.,

$$\Theta \approx T_z^* + \frac{z}{100}. \tag{25}$$

For more accurate computations (when $p_0 \neq 1000$ mb) one can introduce a correction for the deviation of the surface pressure p_0 from 1000 mb. We can find this correction from (6') by substituting $\frac{RT}{p}$ in the latter for v and taking $\Delta p = 1000 - p_0$. In this way we obtain the following formula for the calculation of Θ:

$$\Theta = T_z^* + \frac{z}{100} + \frac{1000 - p_0}{12.5}. \tag{26}$$

Let us see how Θ changes with height in the atmosphere. To do this we take the logarithmic derivative of the expression (24) with respect to z, bearing in mind that in this case the star should be deleted and that we should take partial derivatives:

$$\frac{1}{\Theta}\frac{\partial \Theta}{\partial z} = \frac{1}{T}\frac{\partial T}{\partial z} - \frac{AR}{c_p}\frac{1}{p}\frac{\partial p}{\partial z}. \tag{27}$$

Since

$$\frac{\partial p}{\partial z} = -\rho g = -\frac{pg}{RT},$$

we have

$$\frac{1}{\Theta}\frac{\partial \Theta}{\partial z} = \frac{1}{T}\left(\frac{\partial T}{\partial z} + \frac{Ag}{c_p}\right)$$

or

$$\frac{\partial \Theta}{\partial z} = \frac{\Theta}{T}\left(\frac{\partial T}{\partial z} + \frac{Ag}{c_p}\right). \tag{28}$$

But $\frac{Ag}{c_p} = \gamma_a$ is the dry adiabatic gradient and $\frac{\partial T}{\partial z} = -\gamma$, where γ is the actual gradient of temperature in the atmosphere, and therefore

$$\frac{\partial \Theta}{\partial z} = \frac{\Theta}{T}(\gamma_a - \gamma). \tag{29}$$

Since for average atmospheric conditions γ is usually smaller than γ_a, it follows that $\frac{\partial \Theta}{\partial z} > 0$, i.e., in the atmosphere Θ usually increases with altitude. We note further that the value of the derivative $\frac{\partial \Theta}{\partial z}$ depends not only on the magnitude and sign of the sum $\left(\frac{\partial T}{\partial z} + \frac{Ag}{c_p}\right) = \gamma_a - \gamma$ but also on the value of the ratio $\frac{\Theta}{T}$. Since Θ usually increases with height while T decreases, the ratio $\frac{\Theta}{T}$ increases with height in general and therefore $\frac{\partial \Theta}{\partial z}$ increases in absolute value with height even for a constant $\left(\frac{\partial T}{\partial z} + \gamma_a\right)$. If the sum $\frac{\partial T}{\partial z} + \gamma_a = 0$, i.e., $-\frac{\partial T}{\partial z} = \gamma = \gamma_a$, then $\frac{\partial \Theta}{\partial z} = 0$ and $\Theta = \text{const}$.

From expression (24), after taking its logarithm and differentiating, we find that

$$\frac{d\theta}{\theta} = \frac{dT^*}{T^*} - \frac{AR}{c_p}\frac{dp}{p},$$

or

$$\frac{d\theta}{\theta} = \frac{dT^*}{T^*} - 0.288\frac{dp}{p}. \tag{30}$$

From (13'), in an adiabatic process the right-hand side of the above is zero, i.e.,

$$\frac{d\theta}{\theta} = 0. \tag{31}$$

Thus in adiabatic processes the potential temperature of the air mass does not change whereas the absolute temperature T^* rises with increasing p (descent) or drops with decreasing p (ascent). From (30) it follows that the relative temperature change in such processes is about 29 % of the relative pressure variation.

If the potential temperature of an air mass increases or decreases as it moves vertically, this shows that an outflow or inflow of heat is taking place. Indeed, from equation (4') we have, for a particle,

$$dQ = c_p T^* \left(\frac{dT^*}{T^*} - \frac{AR}{c_p}\frac{dp}{p}\right), \tag{32}$$

and, taking (30) into account,

$$dQ = c_p T^* \frac{d\theta}{\theta}.$$

Using the fundamental equation of hydrostatics $\frac{dp}{p} = -\frac{gdz}{RT}$, one can write expression (30) in the form

$$c_p T^* \frac{d\theta}{\theta} = c_p dT^* + Ag\frac{T^*}{T}dz,$$

or

$$c_p T^* \frac{d\theta}{\theta} = c_p dT^* + Agdz + Ag\frac{T^* - T}{T}dz.$$

Here the first term $c_p dT^*$ gives the change in the heat content of the air mass in question or, as it is called in heat engineering, in its enthalpy $d\varepsilon$; the second term $Agdz$ gives the change in the potential energy dP; and, finally, the third term $Ag\frac{T^* - T}{T}dz$ characterizes the change in the so-called energy of instability of the air mass dE. The last equation may therefore be written as

$$c_p T^* \frac{d\theta}{\theta} = d\varepsilon + dP + dE = d\Pi, \tag{33}$$

where Π is the total energy of an individual air mass. Hence it is obvious that, in the adiabatic displacement of an air mass, since $\frac{d\theta}{\theta} = 0$, $d\Pi$ is also zero and $\Pi = \text{const}$, i.e., its supply of total energy remains constant.

§ 6. Adiabatic changes of state of moist air

Let us extend the conclusions of the previous sections to the case of moist air in which water vapor has not yet reached the state of saturation.

Consider one gram of moist air containing q grams of water vapor. It can be treated as a mixture of $(1-q)$ g of dry air at the partial pressure $(p-e)$ mb and q g of water vapor at the pressure e mb. A quantity of heat dQ supplied to this mixture can be divided into two parts: dQ_1, supplied to dry air, and dQ_2, supplied to water vapor; thus

$$dQ = dQ_1 + dQ_2. \tag{34}$$

Applying the fundamental relations derived above to $(1-q)$ g of dry air and to q g of water vapor, we obtain

$$dQ_1 = c_p(1-q)dT^* - ART^*(1-q)\frac{d(p-e)}{p-e} \tag{35}$$

and

$$dQ_2 = c_{p,\,w}qdT^* - AR_wT^*q\frac{de}{e}, \tag{36}$$

where $c_{p,w}$ and R_w are the specific heat and gas constant for water vapor.

Let the air ascend without exchange of moisture, i.e., under the condition that $q = 0.622\frac{e}{p}$ does not change. From this it follows that

$$\frac{dq}{q} = \frac{de}{e} - \frac{dp}{p} = 0 \quad \text{or} \quad \frac{de}{e} = \frac{dp}{p},$$

and also

$$\frac{d(p-e)}{p-e} = \frac{dp}{p}, \quad \text{since} \quad e \ll p.$$

Taking these relations into account and setting

$$R_w = 1.6R \quad \text{and} \quad c_{p,\,w} = 1.83c_p$$

in (36), we introduce the resulting expression and equation (35) into (34). We then have

$$dQ = dQ_1 + dQ_2 = c_p(1+0.83q)\,dT^* - ART^*(1+0.6q)\frac{dp}{p}. \tag{37}$$

Since for adiabatic processes $dQ=0$, we find that

$$\frac{dT^*}{T^*} = \frac{AR}{c_p}\frac{1+0.6q}{1+0.83q}\frac{dp}{p}. \tag{38}$$

Introducing

$$dp = -\rho g dz = -\frac{p}{RT}g dz,$$

we obtain

$$\frac{dT^*}{dz} = \frac{-Ag}{c_p}\frac{1+0.6q}{1+0.83q}\frac{T^*}{T}. \tag{39}$$

By comparing expressions (21) and (39) one can see that they differ only by the factor

$$\frac{1+0.6q}{1+0.83q}.$$

Since in the atmosphere it is nearly always true that $q < 4\cdot 10^{-2}$ g/g,

$\frac{1+0.6q}{1+0.83q} \approx 1$. Noting also that $T^* \approx T$, instead of (39) we can write

$$\frac{dT^*}{dz} = -\frac{Ag}{c_p} = -\gamma_a, \tag{40}$$

i. e., in practice the adiabatic change in the temperature of moist air (containing water vapor in amounts which do not reach saturation) differs only slightly from the adiabatic change in the temperature of dry air. Consequently, one can assume that for moist air as well

$$-\frac{dT^*}{dz} = \gamma_a \approx 1°/100 \text{ m}.$$

Whereas q remains constant when moist air rises, an important characteristic of air, the relative humidity $f = \frac{e}{E} \cdot 100\%$, does change. This question is discussed in § 10.

§ 7. Moist adiabatic processes. The moist adiabatic gradient

Let us consider the thermodynamic processes which take place in the cooling of an ascending air mass after its temperature has dropped to the dew point τ, which corresponds to the beginning of vapor condensation. Two cases must be distinguished here:

1) the products of condensation, in the form of water droplets and ice crystals, remain within the volume of air under consideration and participate in its movement;

2) the products of condensation precipitate partially or completely from the volume in question.

In the first case the process is reversible (as are dry adiabatic processes) and the system can return to the initial state by going backward; such processes are termed moist adiabatic.

In the second case the process is irreversible and is called pseudo-adiabatic. Here the system cannot be returned to the initial state once it has reached the final one owing to the reduction of its moisture content. Thermodynamic computations for pseudo-adiabatic processes are possible only when the quantity of water precipitated as a result of condensation of water vapor is known.

Let us first examine moist adiabatic processes. We will confine ourselves to a simplified derivation. It follows from the above definition that the total moisture content of the air remains constant in this process and that only the ratio between the gaseous and condensed phase of water is changed.

Let us take one gram of moist air with a specific humidity of q g/g. If dq g of water vapor condenses inside it, a quantity of heat given by Ldq cal/g, where L is the latent heat of vaporization, will be liberated. Since $L \approx 597 - 0.6t$, the quantity of heat liberated can be very significant.

Therefore, instead of (4") we write

$$dQ = c_p dT^* - ART^* \frac{dp^*}{p^*} + Ldq. \tag{41}$$

For an adiabatic process dQ is zero and therefore on the basis of the above

$$dT^* = \frac{AR}{c_p} T^* \frac{dp^*}{p^*} - \frac{L}{c_p} dq. \tag{42}$$

Since the process is quasi-static, $\frac{dp^*}{p^*} = \frac{dp}{p} = -\frac{g}{RT} dz$ (where, as before, T is the temperature of the surrounding medium and p its pressure). Consequently, the temperature change of an air particle containing saturated water vapor will be

$$dT^* = -\frac{Ag}{c_p} \frac{T^*}{T} dz - \frac{L}{c_p} dq. \tag{43}$$

Setting $\gamma_{wa} = -\frac{dT^*}{dz}$ (the so-called moist adiabatic gradient), we can write

$$\gamma_{wa} = -\frac{dT^*}{dz} = \frac{Ag}{c_p} \frac{T^*}{T} + \frac{L}{c_p} \frac{dq}{dz}. \tag{44}$$

It is obvious that for dry air $(dq=0)$ this expression transforms into expression (21) for γ_a.

Since the specific humidity of air in the state of saturation (buoyancy E) is given by

$$q = 0.622 \frac{E}{p},$$

then, taking logarithms and differentiating, we have

$$\frac{1}{q} \frac{dq}{dz} = \frac{1}{E} \frac{dE}{dz} - \frac{1}{p} \frac{dp}{dz},$$

after which, bearing in mind that $\frac{1}{p} \frac{dp}{dz} = -\frac{g}{RT}$, we write

$$\frac{dq}{dz} = q \left[\frac{1}{E} \frac{dE}{dT^*} \frac{dT^*}{dz} + \frac{g}{RT} \right]. \tag{45}$$

We introduce the expression we have obtained into (44); then

$$-\frac{dT^*}{dz} = \frac{Ag}{c_p} \frac{T^*}{T} + \frac{Lq}{c_p} \frac{1}{E} \frac{dE}{dT^*} \frac{dT^*}{dz} + \frac{Lq}{c_p} \frac{g}{RT},$$

or

$$-\frac{dT^*}{dz} \left[1 + \frac{Lq}{c_p E} \frac{dE}{dT^*} \right] = \frac{Ag}{c_p} \frac{T^*}{T} + \frac{Lq}{c_p} \frac{g}{RT}. \tag{46}$$

Assuming, as before, that $T \approx T^*$ and bearing in mind that $q = 0.622 \frac{E}{p}$, we obtain finally

$$-\frac{dT^*}{dz} = \gamma_{wa} = \frac{\frac{Ag}{c_p} \left[1 + \frac{L}{A} \frac{q}{RT} \right]}{1 + \frac{Lq}{c_p E} \frac{dE}{dT^*}} = \frac{Ag}{c_p} \frac{\left[p + 0.622 \frac{LE}{ART} \right]}{\left[p + 0.622 \frac{L}{c_p} \frac{dE}{dT^*} \right]}, \tag{47}$$

or

$$\gamma_{wa} = \frac{Ag}{c_p} \frac{p+a}{p+b} = \gamma_a \frac{p+a}{p+b}, \tag{47'}$$

where

$$a = 0.622 \frac{LE}{ART} \quad \text{and} \quad b = 0.622 \frac{L}{c_p} \frac{dE}{dT^*}.$$

The coefficients a and b have different numerical values for different temperatures and increase rapidly with increasing temperature; further, $a < b$ always, and therefore $p+a<p+b$ and $\frac{p+a}{p+b}<1$, i.e., $\gamma_{wa}<\gamma_a$ always.

Table 17 gives the values of γ_{wa} for certain values of p and T.

TABLE 17

Values of the moist adiabatic gradient (°/100m)

p	$t°$										
(mb)	-60	-50	-40	-30	-20	-10	0	10	20	30	40
1000	1.00	0.99	0.98	0.94	0.88	0.78	0.66	0.54	0.44	0.37	0.31
800	1.00	0.99	0.97	0.93	0.85	0.74	0.62	0.50	0.40	0.34	0.29
600	1.00	0.99	0.96	0.91	0.82	0.69	0.55	0.44	0.36	0.30	0.27
400	0.99	0.98	0.94	0.87	0.75	0.60	0.47	0.38	0.31	0.27	0.24
200	0.98	0.96	0.89	0.76	0.61	0.46	0.35	0.29	0.25	0.22	0.20
100	0.97	0.91	0.80	0.63	0.46	0.34	0.27	0.23	0.20	0.18	0.16

As one can see from the above formulas and from Table 17, γ_{wa}, unlike γ_a, does not remain constant but depends on p and T. The values of γ_{wa} are smallest at high temperatures (when q and E are large) and low pressures. Also, γ_{wa} increases with decreasing temperature and at low temperatures, when the moisture content is low, it approaches γ_a. For purposes of comparison, Figure 27 shows curves giving the march of the

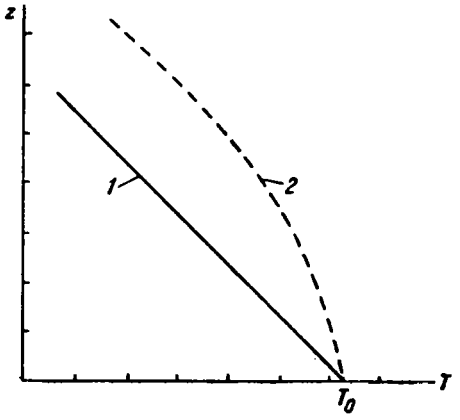


FIGURE 27. Dry (1) and moist (2) adiabatics

adiabatic change in temperature with height of dry air (dry adiabatic) and of moist saturated air ($f = 100\%$) for the same initial temperature — the so-called moist adiabatic.

§ 8. Entropy and potential temperature

In the atmosphere changes of state of systems are often accompanied by the liberation of considerable amounts of heat, as in the condensation of water vapor. In such cases it is possible to have the transformation of internal into mechanical energy. When studying this process it is necessary to take the second law of thermodynamics into account.

From physics textbooks it is well known that for any Carnot cycle, according to Clausius inequality, the sum of the reduced quantities of heat $\frac{\Delta Q}{T}$ cannot be greater than zero.

If we consider any cyclic process then, after breaking it down by means of isothermals and adiabatics into a large number of elementary Carnot cycles, we can write

$$\sum \frac{\Delta Q}{T} \leqslant 0. \tag{48}$$

The equality sign holds for reversible processes, for which the sum can be replaced by the integral

$$\oint \frac{dQ}{T} = 0 \tag{49}$$

(the circle on the integral sign indicates that the latter extends over the entire reversible cycle).

For irreversible processes the sum cannot be replaced by an integral since $\frac{\Delta Q}{T}$ is not a total differential.

In the case of reversible processes the sum of the reduced quantities of heat imparted to the body does not depend on the path along which the process takes place. Hence it follows that on transition from a state A to another state B the integral $\int_A^B \frac{dQ}{T}$ does not depend on the path but is determined exclusively by the initial and final states of the body. It therefore follows that there exists a certain physical quantity φ such that the difference between its values in the states A and B is equal to the sum of the reduced quantities of heat for any reversible process taking place between these states, i.e.,

$$\varphi_B - \varphi_A = \int_A^B \frac{dQ}{T}. \tag{50}$$

The physical quantity φ, which is a function of the state, is called the entropy.

Since $\int_A^B \frac{dQ}{T}$ only gives the change in entropy, the entropy itself of a given state is determined only up to a certain additive constant

$$\varphi = \int \frac{dQ}{T} + \varphi_0, \tag{51}$$

where φ_0 is the entropy of a certain state taken to be the initial one.

For the total differential of the entropy we have, differentiating (51),

$$d\varphi = \frac{dQ}{T}. \tag{52}$$

In meteorology the initial state for the determination of entropy is chosen to be $T_0 = 100°K$ and $p_0 = 1000$ mb.

Making use of expression (4'') for the energy influx

$$dQ = c_p dT - ART \frac{dp}{p},$$

and using (52), we find

$$d\varphi = c_p \frac{dT}{T} - AR \frac{dp}{p}, \tag{53}$$

or

$$\varphi = c_p \ln \frac{T}{T_0} - AR \ln \frac{p}{p_0}. \tag{54}$$

Since, as we stated above, it is assumed by convention that $T_0 = 100°K$ and $p_0 = 1000$ mb, we have

$$\varphi = c_p \ln \frac{T}{100} - AR \ln \frac{p}{1000},$$

or

$$\varphi = c_p \ln T - AR \ln p + \varphi_0. \tag{55}$$

This is how entropy is expressed for dry air.

In view of the fact that the right-hand side of (53), from equation (30), is equal to $c_p \frac{d\theta}{\theta}$, where θ is the potential temperature, we can write it in the form

$$d\varphi = c_p \frac{d\theta}{\theta}, \tag{56}$$

whence

$$\varphi = c_p \ln \theta + \text{const}, \tag{56'}$$

i.e., the entropy of dry air is proportional to the logarithm of the potential temperature.

Since the entropy is constant for reversible adiabatic processes, such processes are often termed isentropic.

The relations derived here are valid for dry and moist air in which the water vapor does not occur in a state of saturation, but become inapplicable when the water vapor reaches saturation and the air, in addition to the water vapor, also contains its condensation products.

In this case, according to (41), the influx of heat is

$$(dQ)_{\text{moist}} = c_p dT - ART \frac{dp}{p} + L dq, \tag{57}$$

where L is the latent heat of vaporization and q the specific humidity. Consequently,

$$(d\varphi)_{\text{moist}} = \frac{(dQ)_{\text{moist}}}{T} = c_p \frac{dT}{T} - AR \frac{dp}{p} + \frac{L}{T} dq. \tag{58}$$

Taking (53) into consideration we obtain

$$(d\varphi)_{\text{moist}} = d\varphi + \frac{L}{T} dq, \tag{59}$$

or

$$\varphi_{\text{moist}} = \varphi_{\text{dry}} + \frac{Lq}{T} + \varphi_0, \tag{60}$$

which shows the very important connection between the entropy of dry air and that of moist air.

§ 9. Dependence of the buoyancy of saturated water vapor on temperature (Clausius-Clapeyron equation)

When the temperature of air decreases the water vapor contained in it ultimately reaches the state of saturation; as the temperature drops further the water vapor condenses and the latent heat of vaporization L is liberated. In the reverse process — rising temperature — the water contained in the air in the liquid state is transformed into vapor.

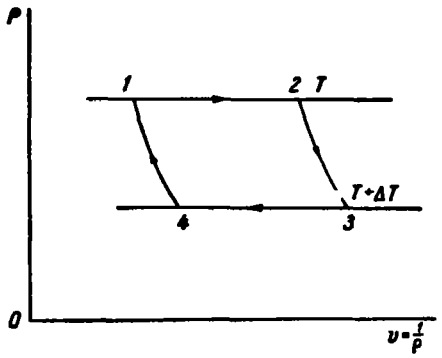


FIGURE 28

In view of the importance of these processes we will examine the system consisting of water and saturated water vapor and study the transformation of water vapor into liquid water and vice versa. Let this system undergo a reversible cyclic process (Figure 28) consisting of the following: 1) isothermal expansion of the system $(1\to2)$ at the temperature T until one gram of water is converted into vapor with the expenditure of the quantity of heat L; 2) adiabatic expansion $(2\to3)$ until the temperature drops to $T-dT$ and the saturation vapor pressure to $E-dE$; 3) isothermal compression $(3\to4)$ resulting in the recondensation of one gram of water; and, finally, 4) adiabatic compression $(4\to1)$, which returns the system to the initial state. The efficiency of the given cycle, which is the ratio of the work done in this reversible process to the quantity of heat supplied, is given by

$$\frac{T-(T-dT)}{T}=\frac{dT}{T}.$$

The work done by the cycle is the sum of the work done in expansion $+E\Delta v$ and the work done in adiabatic contraction $-(E-dE)\Delta v$. The amount of heat delivered to one gram of mass is L and therefore

$$A\,\frac{E\Delta v-[E-dE]\,\Delta v}{L}=A\,\frac{dE\Delta v}{L}=\frac{dE\left(\frac{1}{\rho_{vap}}-\frac{1}{\rho_{liq}}\right)}{L}A. \qquad (61)$$

Since $\frac{1}{\rho_{vap}} = 20.500$ and $\frac{1}{\rho_{liq}} = 1.00013$, we disregard the quantity $\frac{1}{\rho_{liq}}$ in comparison with $\frac{1}{\rho_{vap}}$ and write

$$\frac{dE}{L} = \frac{\rho_{vap}}{A}\frac{dT}{T},$$

whence

$$\frac{dE}{dT} = \frac{L\rho_{vap}}{AT}. \tag{62}$$

But

$$\rho_{vap} = \frac{E}{R_{vap}T},$$

consequently,

$$\frac{dE}{dT} = \frac{LE}{AR_{vap}T^2},$$

or

$$\frac{dE}{E} = \frac{L}{AR_{vap}}\frac{dT}{T^2}. \tag{63}$$

This is the Clausius-Clapeyron equation, familiar from physics, which expresses the temperature dependence of the change in buoyancy of saturated water vapor. We will be using it repeatedly in the following.

§ 10. Condensation level

In an ascending mass of moist air the transformation of the temperature variation from the dry adiabatic to the moist adiabatic takes place at the height at which the water vapor reaches the state of saturation. The question of determining this height, or condensation level, is an important one. In practice the condensation level is near the lower boundary of the clouds.

Let us denote the dew point by τ. If the condensation level is h_c, we have $T^*(h_c) = \tau^*(h_c)$. As the air mass ascends τ changes with the variation of the vapor pressure in the ascending air. Before condensation occurs, however, neither the specific humidity nor the mixture ratio change.

Let us introduce the temperature gradient of the rising mass $\frac{dT^*}{dz}$ and the temperature gradient of the dew point $\frac{d\tau}{dz}$. We can write the following equation for the height of the condensation level:

$$T_0^* + h_c\frac{dT^*}{dz} = \tau_0 + h_c\frac{d\tau}{dz}, \tag{64}$$

from which we find

$$h_c = \frac{T_0^* - \tau_0}{\frac{d\tau}{dz} - \frac{dT^*}{dz}}. \tag{65}$$

Before the condensation level $\frac{dT^*}{dz} = -\gamma_a$. To determine $\frac{d\tau}{dz}$ we will make use of the fact that the maximum vapor pressure at the dew point τ should

be equal to the actual vapor pressure e, i.e., $E=e$, while the specific humidity

$$q(z)=0.622\frac{E}{p} \tag{66}$$

should remain constant to heights $<h_c$.

Taking the logarithm of the above equation and differentiating with respect to z, we obtain

$$\frac{1}{E}\frac{dE}{d\tau}\frac{d\tau}{dz}=\frac{1}{p}\frac{dp}{dz}. \tag{67}$$

Since from the hydrostatic equation $\frac{1}{p}\frac{dp}{dz}=-\frac{g}{RT}$ (where T is the temperature of the surrounding medium), we obtain from (67) that

$$\frac{d\tau}{dz}=-\frac{g}{RT}\frac{1}{\frac{1}{E}\frac{dE}{d\tau}}. \tag{68}$$

But, according to the Clausius-Clapeyron equation,

$$\frac{1}{E}\frac{dE}{d\tau}=\frac{L}{AR_{vap}\tau^2}$$

(cf. Chapter 2).

In view of the fact that $R_{vap}=1.6R$, we have, finally,

$$\frac{d\tau}{dz}=-\frac{g}{RT}\frac{AR_{vap}}{L}=-1.6\frac{Ag\tau^2}{TL}. \tag{69}$$

For approximate computations one can take on the average $g=980\,\text{cm/sec}^2$, $\tau=T=280°$, $L\sim 600\,\text{cal/g}$ and $A=0.24\cdot 10^{-7}\,\text{cal/erg}$. Then

$$\frac{d\tau}{dz}=-\frac{1.6\cdot 980\cdot 280\cdot 0.24\cdot 10^{-7}}{600}\,°/\text{cm}\approx -0.17°/100\,\text{m}.$$

Consequently, on the basis of (65),

$$h_c=\frac{T_0-\tau_0}{1-0.17}100=121\,(T_0-\tau_0)\ \text{m}. \tag{70}$$

In this connection, let us consider the variation of the relative humidity in the adiabatic ascent or descent of air. Taking the logarithm of the relation $f=\frac{e}{E}100\,\%$ and differentiating with respect to z, we find

$$\frac{1}{f}\frac{df}{dz}=\frac{1}{e}\frac{de}{dz}-\frac{1}{E}\frac{dE}{dT^*}\cdot\frac{dT^*}{dz}.$$

Since the specific humidity $q=0.622\frac{e}{p}$ remains constant in this process, whence $\frac{1}{e}\frac{de}{dz}=\frac{1}{p}\frac{dp}{dz}$, we have

$$\frac{1}{f}\frac{df}{dz}=\frac{1}{p}\frac{dp}{dz}-\frac{1}{E}\frac{dE}{dT^*}\frac{dT^*}{dz}.$$

Bearing in mind that for $z=h_c$, $T^*=\tau$

$$\frac{1}{E}\frac{dE}{dT^*}\frac{dT^*}{dz}=\frac{1}{E}\frac{dE}{d\tau}\frac{d\tau}{dz},$$

we can write

$$\frac{1}{f}\frac{df}{dz}=\frac{1}{p}\frac{dp}{dz}\left[1-\frac{\frac{dT^*}{dz}}{\frac{d\tau}{dz}}\right]=\frac{1}{p}\frac{dp}{dz}\left[1+\frac{\gamma_a}{\frac{d\tau}{dz}}\right]. \qquad [71]$$

As an approximate quantitative estimate we will assume that $\frac{d\tau}{dz}=$ $=-0.17°/100\,\text{m}$ and that for the lower atmospheric layers we can take $\frac{1}{p}\frac{dp}{dz}\approx-\frac{1}{1000}0.125$. Then

$$\frac{1}{f}\frac{df}{dz}=-\frac{0.125}{1000}\left[1-\frac{1}{0,17}\right]\approx-\frac{0.125}{1000}(1-5.9)=$$

$$=\frac{0.125\cdot 4.9}{1000}=0.06\cdot 10^{-3}\ \text{m}^{-1}, \qquad (72)$$

i.e., the relative humidity of adiabatically rising moist air increases by approximately 6% of its surface value for every 100 m of ascent.

§ 11. Equivalent and equivalent-potential temperature. Pseudo-adiabatic processes

In adiabatic processes taking place in dry or moist ("unsaturated") air the potential temperature Θ, as was shown in § 5, remains constant and is a convenient characterization of the state of the air. The situation is different when the air contains saturated water vapor and condensation products. In this case, since the temperature variation occurs with the gradient $\gamma_{wa}<\gamma_a$, the potential temperature increases with height and there is no advantage in using it. Therefore, use is made of other temperature characterizations in which the latent heat of vaporization is accounted for beforehand. The simplest of these is the equivalent temperature T_e and equivalent-potential temperature Θ_e.

Let us consider one gram of moist air containing q g of water vapor at the temperature T. If all the water vapor were to condense completely, Lq calories of heat would be given off and the temperature of the air would be raised by $\Delta T=\frac{Lq}{c_p}$. The temperature of the air would thus be

$$T_e=T+\Delta T=T+\frac{Lq}{c_p}. \qquad (73)$$

Taking $L=597\,\text{cal/g}$ for $t=0°$, $c_p=0.24\,\text{cal/g}\cdot\text{deg}$ and $\frac{L}{c_p}\approx 2500$, we obtain

$$T_e=T+2500q \qquad (74)$$

(if q is expressed in g/g).

Since $q=0.622\,\frac{e}{p}$ g/g,

$$T_e=T+\frac{1570e}{p}, \qquad (75)$$

which gives, for $p=1000$ mb,

$$T_e=T+1.57e. \qquad (75')$$

For the conversion of vapor into ice one should allow for the latent heat of fusion as well and take $L_{tot} = (597 + 80)$ cal/g in calculations.

The temperature T_e calculated in the manner indicated above is called the equivalent temperature. It is a certain conventional temperature of moist air; more specifically, it is the temperature which the moist air will attain if all the water vapor contained in it condenses and the heat liberated in the process is completely used in heating the air (for $p = \text{const}$). Obviously, T_e does not change during processes of condensation of the water vapor contained in the air. This is the advantage of using it for various calculations.

The quantity $\Delta T = \frac{Lq}{c_p}$ which must be added to be the absolute temperature T of moist air in order to obtain its equivalent temperature is called the equivalent difference.

If air, at the initial temperature T_e, is brought adiabatically (along the dry adiabatic) to the standard pressure $p = 1000$ mb, its temperature will be

$$\Theta_e = T_e\left(\frac{1000}{p}\right)^{0.288} = \left(T + \frac{Lq}{c_p}\right)\left(\frac{1000}{p}\right)^{0.288}, \qquad (76)$$

which is called the equivalent-potential temperature

In moist adiabatic processes Θ_e remains constant.

Let us now consider pseudo-adiabatic processes where part of the condensation products falls out of the air mass under consideration. In moist adiabatic processes in which all the condensed moisture remains inside the air particle the curve of state of the particle is the moist adiabatic; on the other hand, in pseudo-adiabatic processes in which all the moisture is precipitated the curve of state of the particle is the pseudo-adiabatic. There is no difference between the moist adiabatic and pseudo-adiabatic during the particle ascent. On descent in the first (reversible) process the change in particle temperature takes place along the same moist adiabatic as in the ascent; in the second (irreversible) process, owing to precipitation of all the moisture the temperature changes along the dry adiabatic. Therefore in pseudo-adiabatic processes the particle is returned to the original level at a higher temperature than it initially had.

To account for the latent heat of condensation in pseudo-adiabatic processes, the pseudo-equivalent temperature T_{ps} and pseudo-potential temperature Θ_{ps} are introduced.

The pseudo-equivalent temperature T_{ps} is the temperature which the particle will attain if it is raised to the condensation level along the dry adiabatic and above this level along the moist adiabatic until all the water vapor contained inside it condenses, and then lowered along the dry adiabatic to the initial level. In practice the effect of condensation is accounted for by determining the equivalent difference at the pressure corresponding to the condensation level.

The pseudo-potential temperature Θ_{ps} is the temperature which the particle will attain if after being raised pseudo-adiabatically until complete condensation of the water vapor it is then lowered dry-adiabatically to the level $p = 1000$ mb. Obviously, if the displacement of a moist particle takes place adiabatically, its pseudo-adiabatic temperature will remain constant.

We note further that another characteristic of moist adiabatic processes is also used occasionally. This is the so-called potential wet-bulb temperature Θ', by which is meant the temperature which a particle

will attain if it is lowered moist-adiabatically from the condensation level to the level $p = 1000$ mb. Like Θ_{ps}, Θ' remains unchanged in vertical displacements of moist air particles.

The following relations hold between these conventional temperatures:

$$T < T_e < T_{ps}, \tag{77}$$

and

$$\Theta < \Theta_e < \Theta_{ps}. \tag{78}$$

The calculation of Θ_e and Θ_{ps} is usually carried out by graphical methods using the so-called adiabatic diagrams (cf. § 13).

§ 12. Conditions of vertical stability of the atmosphere

Vertical motions are constantly forming in the atmosphere. They play a very large role in many processes, notably the transport of heat and moisture and the formation of clouds and precipitation. The appearance of these motions is due to a number of causes, the most important of which are: the presence of a temperature difference between a certain air particle and the surrounding medium, the incursion of air streams over various irregularities of the ground, and so on. The paramount factor for the later development and intensity of an initially vertical motion is the ratio between the temperature of the vertically moving air mass $T^*(z)$ and the temperature of the surrounding medium $T(z)$, or, in other words, the vertical distribution of temperature in the atmosphere.

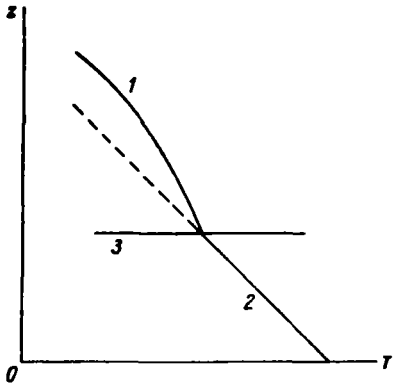


FIGURE 29. 1-moist adiabatic; 2-dry adiabatic; 3-condensation level

In evaluating the ratio of $T^*(z)$ to $T(z)$ it is convenient to employ a graphical method. The curve plotted in the coordinate system (T, z) and indicating the change in temperature $T^*(z)$ of the ascending air particle as a function of the height z is called the curve of state. In the case of moist air it consists of the dry adiabatic to the height of the condensation level (i.e., that height at which condensation of the water vapor begins), then of the moist adiabatic (Figure 29).

The curve which represents the actual distribution of temperature in the atmosphere $T(z)$ is called the stratification curve. It can only be constructed on the basis of actual measurements of the temperature at different levels.

That state of the atmosphere — or its temperature stratification, as it is called — in which vertical movements tend to grow (i.e., vertically displaced air masses experience positive accelerations, i.e., along the direction of the velocity) is called the unstable state. If, on the other hand, vertical movements do not grow but are attenuated (air masses

experience negative accelerations, i.e., in the direction opposite to the velocity), one then speaks of the stable state. Finally, if particles are not accelerated during vertical displacement, neutral equilibrium is said to prevail.

Let us first consider the conditions of stability of dry air.

From the hydrostatic equation it follows that air will be in equilibrium when the vertical component of the force of hydrostatic pressure, referred to the unit mass and given by $\frac{1}{\rho}\frac{dp}{dz}$, is balanced by the force of gravity g. This holds if horizontal uniformity of temperature and quasi-stationarity are assumed. However, for various reasons it may happen that $T^* \neq T$ and $\rho^* \neq \rho$; then the air mass in question, under the influence of the force

$$f = g\rho - g\rho^* \tag{79}$$

is accelerated along the vertical and begins to move. The magnitude of this acceleration, i.e., of the force per unit mass, is expressed according to Archimedes law by

$$\frac{d^2z}{dt^2} = \frac{g(\rho - \rho^*)}{\rho^*} = g\left(\frac{\rho}{\rho^*} - 1\right), \tag{80}$$

or, since $\frac{\rho}{\rho^*} = \frac{T^*}{T}$ (under quasi-stationarity condition, where $p = p^*$),

$$\frac{d^2z}{dt^2} = g\left(\frac{T^*}{T} - 1\right) = g\frac{T^* - T}{T}. \tag{81}$$

Hence it follows that if an air mass is warmer than the surrounding air at a certain level $(T^* > T)$ it experiences a positive acceleration and is lifted upward; for $T^* < T$ it experiences negative acceleration and is dropped downward. The magnitude of this acceleration may be insignificant. Indeed, if the temperature of the surrounding air is $T = 273°K$ it is $3.6\ cm/sec^2$ for a difference $T^* - T = 1°$.

In such displacements the temperature of an air mass under adiabatic conditions will change according to the law

$$T^* = T_0^* - \gamma_a \Delta z, \tag{82}$$

where T_0^* is the initial temperature and Δz the distance along which the particle is moved.

Let us assume that the temperature of the surrounding air also changes linearly, i.e.,

$$T = T_0 - \gamma \Delta z, \tag{82'}$$

where γ is the actual vertical temperature gradient.

Then

$$T^* - T = (T_0^* - T_0) - (\gamma_a - \gamma)\Delta z \tag{83}$$

and, consequently, the acceleration (81) will be

$$\frac{d^2z}{dt^2} = g\frac{T_0^* - T_0}{T} - g\frac{\gamma_a - \gamma}{T}\Delta z. \tag{84}$$

Hence the acceleration is determined by two factors the effects of which are represented by the two terms in the right-hand side of the above. The first term expresses the dependence of the acceleration on the value of the initial heat excess of the particle $(T_0^* - T_0)$ relative to the surrounding air.

The second gives the dependence of the acceleration on the stratification of the atmosphere $(\gamma_a - \gamma)$.

Let us write (83) in the form

$$\Delta T = \Delta T_0 - (\gamma_a - \gamma)\Delta z.$$

From this we can easily find the height $\Delta z = h$ by which the particle must be displaced in order to arrive at the temperature of the surrounding air, i.e., in order for $T^* - T = \Delta T = 0$. This height will be

$$\Delta z = h = \frac{\Delta T_0}{\gamma_a - \gamma} = \frac{T_0^* - T_0}{\gamma_a - \gamma}. \tag{85}$$

The height h is called the convection level and, as one can see from (85), depends on the initial temperature difference $(T_0^* - T_0)$ and on the actual vertical temperature gradient γ.

The initial vertical displacement of an air mass can be brought about not only by heating or cooling but also by mechanical factors. If a certain air mass is displaced along the vertical and its temperature T^* becomes different from that of the surrounding medium, the possibility of further motion (after the action of the external forces terminates) is determined by the difference $(T^* - T)$ and the magnitude of the gradient γ. Indeed, in this case (as $\Delta T_0 = T_0^* - T_0 = 0$), we obtain

$$T^* - T = \Delta T = -(\gamma_a - \gamma)\Delta z,$$

and the acceleration of the particle is

$$\frac{d^2z}{dt^2} = \frac{g}{T}(T^* - T) = -\frac{g}{T}(\gamma_a - \gamma)\Delta z. \tag{86}$$

If $\gamma_a - \gamma = 0$, i.e., $\gamma_a = \gamma$, the acceleration is zero and the air mass, having been lifted to a certain height by external forces, remains at this height when the latter cease to act; this corresponds to neutral equilibrium. If $\gamma_a - \gamma > 0$, i.e., $\gamma < \gamma_a$, the acceleration is negative and the air mass returns to its initial position when the action of the external forces ceases. This is the so-called stable equilibrium. Finally, if $\gamma_a - \gamma < 0$, i.e., $\gamma > \gamma_a$, the acceleration is positive and an air mass removed from its initial position will continue to move. This corresponds to unstable equilibrium.

Making use of relation (29) for the variation of the potential temperature Θ with height

$$\frac{\partial \Theta}{\partial z} = \frac{\Theta}{T}(\gamma_a - \gamma),$$

expression (86) for the vertical acceleration of a displaced mass can also be written in the form

$$\frac{d^2z}{dt^2} = -\frac{g}{\Theta}\frac{\partial \Theta}{\partial z}\Delta z. \tag{87}$$

From the above it follows that in the state of neutral equilibrium, in which $\frac{d^2z}{dt^2} = 0$, Θ does not vary with height $\left(\frac{\partial \Theta}{\partial z} = 0\right)$; in the stable state $\left(\frac{d^2z}{dt^2} < 0\right)$ Θ increases with height $\left(\frac{\partial \Theta}{\partial z} > 0\right)$; finally, in the unstable state $\left(\frac{d^2z}{dt^2} > 0\right)$ Θ decreases with height $\left(\frac{\partial \Theta}{\partial z} < 0\right)$. We note that in (87) Θ refers to the surrounding medium.

Summing up, the state of the atmosphere is stable if $\gamma < \gamma_a$ or if Θ increases with altitude; unstable if $\gamma > \gamma_a$ or if Θ decreases with altitude; neutral if $\gamma = \gamma_a$ or if Θ does not vary with height.

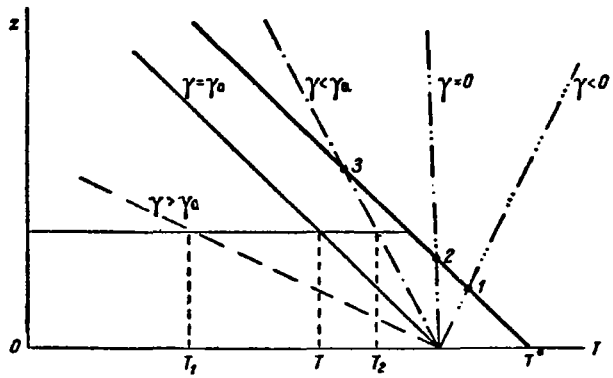


FIGURE 30. Conditions of stability for dry air. The points 1, 2, 3 indicate the height to which the warmed particle T is raised for different temperature gradients γ in the surrounding atmosphere

This statement is represented graphically in Figure 30. The figure shows the height of ascent of an air mass which is warmer than the surrounding air at the initial level. The points of intersection between the curve of state (heavy line) and the curves of stratification of the atmosphere (broken lines) give the height of the convection level h: the point 1 for the case where the temperature in the atmosphere increases linearly with height (a so-called inversion is present, $\gamma < 0$), the point 2 for the case where temperature does not change with height (isothermy, $\gamma = 0$) and the point 3 for the case where, as is usual, temperature decreases with height ($\gamma > 0$). The figure also shows that when temperature in the atmosphere decreases with height by more than 1°/100 m ($\gamma > \gamma_a$, i.e., the state is unstable), the height h cannot be specified.

This method of analyzing the stability of the atmosphere allows one to explore the conditions of development of convection. As it is based on the examination of temperature variations in an individual particle (air mass) during vertical displacement, it is known as the particle method.

However, an essential drawback of this method is that it fails to account for opposing compensatory movements of air. Consequently, other methods of determining stability have been put forward (see § 15).

It should also be recalled that in actual conditions some mixing and heat exchange between the moving air mass and the surrounding medium always take place. In this case the processes can no longer be considered as adiabatic. With regard to the conditions of applicability of the relations derived above, it is found that the smaller the influx of heat to the particle and the greater its vertical velocity of motion, the closer the processes are to being adiabatic.

We shall now investigate the conditions of vertical stability for moist air.

The conditions of development of vertical motions in the atmosphere for the case of moist air can be obtained by comparing the curve of the

actual temperature distribution (stratification curve) with the curve of state for moist air (i. e., with the dry adiabatic up to the condensation level and with the moist adiabatic above this level). As before, the state will be stable if $\gamma<\gamma_{wa}$, unstable if $\gamma>\gamma_{wa}$, and neutral if $\gamma=\gamma_{wa}$.

The possible relationships between the dry and moist adiabatics and the stratification curve (curves 1, 2 and 3) are represented graphically in Figure 31.

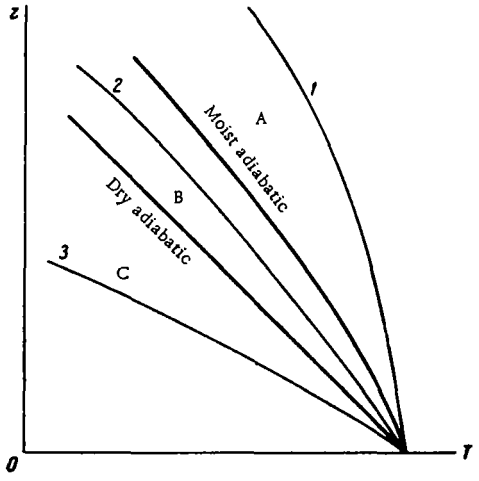


FIGURE 31. Conditions of stability for moist air

1, 2, 3–stratification curves; A–stable-state region; B–region of dry-stable but moist-unstable state; C–unstable-state region.

Curve 1 which lies in region *A* to the right of the moist adiabatic corresponds to the absolutely stable state, for which $\gamma<\gamma_{wa}<\gamma_a$. Curve 2, lying in region B between the dry and the moist adiabatic, characterizes the stable state for dry air ($\gamma<\gamma_a$) but the unstable state for moist air—the moist-unstable state—for which $\gamma>\gamma_{wa}$. Finally, curve 3 lying in region C to the left of the dry adiabatic characterizes the unstable state for both dry and moist air ($\gamma>\gamma_{wa}>\gamma_a$). In cases where $\gamma=\gamma_a$ we have the neutral state for dry air and for $\gamma=\gamma_{wa}$ the neutral state for moist air.

Thus from the standpoint of dry adiabatic and moist adiabatic processes the following types of atmospheric stratification exist:

1) $\gamma>\gamma_a>\gamma_{wa}$ — dry- and moist-unstable, or absolutely unstable, stratification;

2) $\gamma=\gamma_a>\gamma_{wa}$ —dry-neutral and moist-unstable stratification;

3) $\gamma_a>\gamma>\gamma_{wa}$ —dry-stable and moist-unstable, or conditionally stable, stratification;

4) $\gamma_a>\gamma=\gamma_{wa}$ —dry-stable and moist-neutral stratification;

5) $\gamma<\gamma_{wa}<\gamma_a$ —dry- and moist-stable, or absolutely stable, stratification.

A very important question is that of the stability of atmospheric layers (large volumes) displaced along the vertical. Processes of this kind

develop, for instance, when air masses are lifted along frontal surfaces or moved vertically (Figure 32a and b). We will assume that initially the layer was absolutely stable. In both figures the line AB represents the temperature distribution in the layer before lifting.

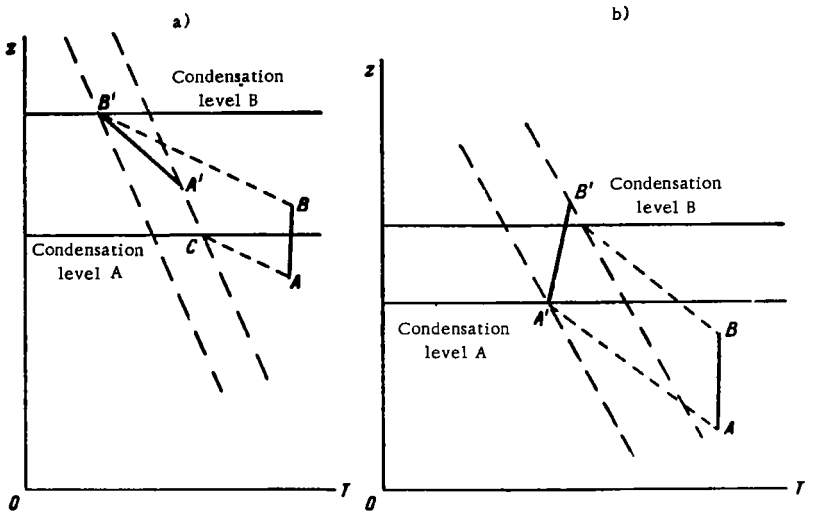


FIGURE 32

a–convective instability; b–convective stability.

If the relative humidity in the lower portion of the layer (point A) is substantially higher than in its upper portion (point B; see Figure 32a), a particle lifted from the point A along the dry adiabatic AC will rapidly attain the condensation level (point C). On further lifting along the moist adiabatic CA' it will be cooled comparatively slowly, and at the point A' it will attain a certain temperature T'. On the other hand, a particle lifted from the point B at a lower humidity will rise along the dry adiabatic BB' for a longer time and be cooled comparatively strongly. As a result, the stratification of the layer after lifting, characterized by the curve $A'B'$, will be moist-unstable ($\gamma > \gamma_{wa}$), and the layer as a whole, since it will then be saturated, will also be stratified unstably.

By using similar reasoning we arrive at the conclusion that if the humidity of a layer increases with height (Figure 32b) its stratification will become even more stable on adiabatic lifting. In the former case the stratification of the layer is termed convectively unstable, in the latter case convectively stable. The type of stability can be determined from the state of the air particles lying at the lower and upper boundaries of the layer. This question is of great significance for evaluating the development of clouds and convective movements inside them.

§ 13. Thermodynamic graphs (diagrams)

In many thermodynamic computations, especially where speed as well as graphic representation of the results are required, a variety of so-called thermodynamic graphs are employed. We will consider the most current of these, the aerological diagram. This diagram consists of a graph sheet with the temperature (from −65 to +40°) plotted along the abscissa and the logarithm of the pressure (two scales for p from 1050 to 200 mb and from 250 to 50 mb) along the ordinate axis. Isobars are indicated at 10 mb intervals by horizontal straight lines and isotherms at one-degree intervals by vertical straight lines. The following curves are plotted on the resulting grid:

a) dry adiabatic (inclined lines);

b) moist adiabatic (broken blue lines);

c) specific humidity q_{max} for the state of saturation (solid blue lines; blue numbers near these lines give the values of q_{max}).

Short blue vertical dashes on the isobaric lines (1000, 900, 800 mb, etc.) indicate the difference between the virtual and the true temperature for the state of saturation. In addition, blue numbers on the ordinate axis indicate the difference in dynamical meters between the levels of the principal isobaric surfaces (1000, 900, 800 mb, etc.) at a temperature of 0°.

On the blank aerological diagram are plotted the curve of temperature distribution based on aerological data, i.e., the stratification $T(p)$, and also the curve of state $T^*(p)$ for an ascending air mass. This curve is constructed from values of the temperature, pressure and relative humidity at the initial (zero) level. It consists of the dry adiabatic to the condensation level and of the moist adiabatic above it.

. With the help of this diagram it is possible to determine a number of crucial thermodynamical characteristics. For example, the height of the condensation level is located on the diagram by means of the q_{max} lines and is determined by the point of intersection between the dry adiabatic and the isogram of the initial value of q. One can also determine the values of the actual specific humidity from the temperature, pressure and relative humidity (q_{max} is read from the diagram and the actual value of q found from the value of the relative humidity). The dew point, potential, pseudo-equivalent and pseudo-potential temperatures can also be determined. A particularly important use of the aerological diagram is in the determination of the energy of instability, which will be considered in the following section.

In addition to the diagram described here, other thermodynamic graphs proposed by various authors have found practical application. Examples are the tephigram, with the temperature T and entropy φ as coordinates, the aerogram, with the axes log T and T log p, and Laikhtman's sondogram, with the axes log T and T log Θ.

All such graphs show isobars, isotherms, dry and moist adiabatics, and isograms of specific humidity in the state of saturation q_{max}; their coordinates, however, vary. The same calculations can be carried out with their aid as with the aerological diagram.

Recently L. T. Matveev has elaborated a general method for constructing adiabatic diagrams which allow the determination of the energy of instability while simultaneously satisfying additional conditions dictated by the nature of their application.

§ 14. Energy of instability and its determination

In earlier sections it was shown that the greater the temperature difference between an ascending particle (air mass) and the surrounding medium, the greater the acceleration experienced by the particle. However, it is difficult to characterize the vertical stability of a large layer by means of the vertical gradient, which changes with height. Moreover, the vertical temperature gradient is not a measure of energy. It characterizes the kinetic energy of vertical movements only in a qualitative manner.

Consequently, to characterize the energy stores of an unstable layer of air one uses the concept of the so-called energy of instability.

By energy of instability is understood the energy liberated or expended in the displacement of a unit mass of air from one level (where the pressure is p_0) to another (pressure p); in other words, it is the work done by the buoyancy (Archimedes force) in the vertical lifting of a unit mass.

Since the acceleration of a vertically moving air particle is

$$\frac{dz^2}{dt^2} = g\frac{T^* - T}{T},$$

the work done by a unit air mass upon a displacement dz will be

$$dE = g\frac{T^* - T}{T}dz, \tag{88}$$

and that along the path from z_1 to z_2

$$E = \int_{z_1}^{z_2} g\frac{T^* - T}{T}dz. \tag{89}$$

This work, or more correctly the equivalent energy, is converted into the kinetic energy of the rising air mass.

Since $dp = -g\rho dz$, whence $dz = -\frac{dp}{\rho g} = -\frac{RT}{g}d\ln p$, instead of (89) we obtain

$$E = -R\int_{p_1}^{p_2}(T^* - T)d\ln p = R\int_{p_2}^{p_1}(T^* - T)d\ln p. \tag{90}$$

As one can see from this formula, the energy of instability can be both greater and less than zero depending on the sign of the temperature difference $T^* - T$. If $T^* - T > 0$, i.e., the particle is warmer than the surrounding medium, the energy of instability $E > 0$ and the particle, experiencing a positive acceleration, can do the work without the intervention of external forces. If, however, $T^* - T < 0$, then the energy of instability will be negative and an influx of external energy is required to lift the particle. It is assumed in this argument that the vertical displacement of the air mass takes place adiabatically.

From (90) it is evident that the work done in moving the particle from the initial level z_1 at the pressure p_1 to the level z_2 at the pressure p_2 is represented by the area enclosed between the two isobars p_1 and p_2, the curve of state and the curve of stratification.

The three different cases which can be encountered in the atmosphere are shown graphically in Figure 33. Figure 33a represents the case where at all levels the curve of state lies to the right of the stratification curve. Here $T^* > T$ throughout and therefore the energy of instability (shaded area

on the figure) is positive. This makes it possible for vertical motions to develop in the atmosphere, leading usually to the formation of convective clouds. In the second case (Figure 33b) the curve of state lies to the left of

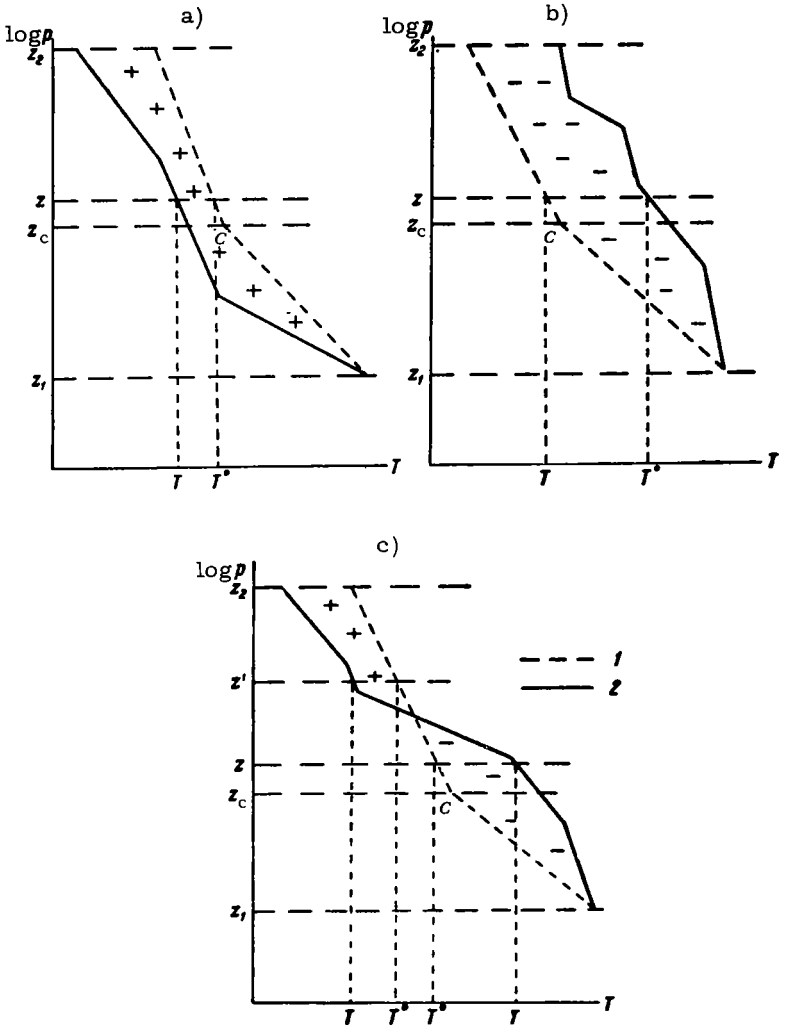


FIGURE 33. Energy of instability for various types of atmospheric stratification

1-curve of state; 2-curve of stratification; z_c-condensation level, T and T^*-temperature of surrounding air and particle.

the stratification curve at all levels. Here, obviously, since $T^* < T$ everywhere, the energy of instability is negative and the upward displacement of particles is impossible. Finally, in the third, more general case (Figure 33c) the curve of state lies to the right of the stratification curve in the upper portion of the layer and to the left of it in the lower portion. This

corresponds to positive (top) and negative (bottom) energies of instability. The overall energy store of the entire layer is the algebraic sum of the energy of instability in the individual sections of the layer.

Returning to equation (88), let us express it in the form

$$dE = \frac{d^2z}{dt^2} dz. \tag{91}$$

Since $\frac{dz}{dt} = w$, where w is the vertical velocity, we can rewrite the above as

$$dE = \frac{dw}{dt} dz = w dw, \tag{92}$$

or

$$dE = d\left(\frac{w^2}{2}\right). \tag{93}$$

Integrating this expression from a certain initial level z_0 where $w = w_0$ to the level z where the velocity is w, we obtain

$$w^2 - w_0^2 = 2E, \quad \text{or} \quad w = \sqrt{2E - w_0^2}, \tag{94}$$

where E is the energy of instability in the entire layer from z_0 to z.

Assuming that at the initial level (e.g., the earth's surface) $w_0 = 0$, we find

$$w = \sqrt{2E}. \tag{95}$$

This formula makes it possible to compute the values of the vertical velocity ω for different values of the energy of instability. From the computed values cited in Table 18 one can see that for large values of E the vertical velocity of motion w can reach very large values amounting to several tens of meters per second.

TABLE 18

Values of the vertical velocity for given energy of instability

E (cm^2/sec^2)	10^2	10^3	10^4	10^5	10^6	10^7
w (m/sec)	0,14	0,45	1,41	4,5	14	45

§ 15. Determination of the instability by the layer method

In § 12 it was mentioned that the particle method of determining atmospheric instability had a number of defects. Thus in analyzing stability by this method one assumes that the particle is moving in a stationary medium, and any influence of the motion of the particle on the surrounding atmosphere is disregarded. In actual conditions when a particle moves in any one direction compensatory motions are formed in the opposite direction. Other methods which account for the omissions of the particle method have therefore been proposed. We will briefly set forth the essential points of one of these, namely the layer method. The foundations of this method were laid by Bjerknes and its practical applications were developed in recent years by N. S. Shishkin.

Let us consider a certain layer z_1, z_2, in which ascending and descending movements are present. We denote the total cross-section of the former by s and their vertical velocity by w; correspondingly we denote the cross-section and velocity of the descending currents by s' and w We will assume that: a) all changes in the layer under consideration take place adiabatically; b) meteorological elements in the layer do not change under the influence of horizontal motions (if such are present); and c) the mass of the entire layer in question remains unchanged. Disregarding small density changes, one can write the last condition (conservation of mass in the layer) in the form

$$sw + s'w' = 0. \tag{96}$$

If at the initial instant $t=0$ the temperature is T_1 at the lower level z_1, T_2 at the upper level z_2 and T at a certain level z inside the layer and the temperature gradient throughout the layer is γ, then obviously $T_1 = T^* + \gamma(z - z_1)$, $T_2 = T + \gamma(z_2 - z)$ and the temperature difference

$$T_1 - T_2 = \gamma(z_2 - z_1). \tag{97}$$

After a certain interval Δt particles from the level z_1 will reach the level z (inside the layer) in ascending currents; their temperature will be $T_1 - \beta w \Delta t$, where β is the adiabatic gradient in the ascending currents. In the descending currents particles from the level z_2 reaching the level z will be at the temperature $T_2 - \beta' w' \Delta t$, where β' is the gradient in the descending currents. Obviously, the temperature difference between these currents under the conditions (b) and (c) is given, from (96) and (97), by

$$\Delta T = (T_1 - \beta w \Delta t) - (T_2 - \beta' w' \Delta t) = \gamma(w - w')\Delta t - (\beta w - \beta' w')\Delta t =$$
$$= \left[(\gamma - \beta) + \frac{w'}{w}(\beta' - \gamma)\right] w \Delta t = \left[(\gamma - \beta) - \frac{s}{s'}(\beta' - \gamma)\right] w \Delta t. \tag{98}$$

If condensation does not take place in the ascending and descending currents, then, according to condition (a), $\beta = \beta' = \gamma_a$ and

$$\Delta T = (\gamma - \gamma_a)\left(1 + \frac{s}{s'}\right) w \Delta t. \tag{99}$$

If on the other hand condensation takes place in the ascending currents and the descending ones are dry adiabatic, then $\beta = \gamma_{wa}$ while $\beta' = \gamma_a$.

Let us assume that the stratification is conditionally stable, i. e., $\gamma_{wa} < \gamma < \gamma_a$. For this case

$$\Delta T = \left[(\gamma - \gamma_{wa}) - \frac{s}{s'}(\gamma_a - \gamma)\right] w \Delta t = (\gamma_a - \gamma)\left[\frac{\gamma - \gamma_{wa}}{\gamma_a - \gamma} - \frac{s}{s'}\right] w \Delta t \tag{100}$$

and it is obvious that ΔT is smaller than it would be in the case of the lifting of a very small particle ($s \approx 0$). Such a layer might be termed selectively unstable, since it is unstable for $\Delta T > 0$, which happens only for $\frac{s}{s'} < \frac{\gamma - \gamma_{wa}}{\gamma_a - \gamma}$. This is possible for sufficiently small dimensions of the ascending currents (small s) or for vertical pulses covering a very small area. On the other hand, the layer is stable when $\Delta T < 0$, i.e., if $\frac{s}{s'} > \frac{\gamma - \gamma_{wa}}{\gamma_a - \gamma}$.

More detailed study of stability conditions shows that the development of convection is determined by the ratio between the masses of saturated and unsaturated air.

If the rising and falling masses are unsaturated (or, on the contrary, saturated), conditions of atmospheric stability will not depend on the ratio between these masses. In such cases the conditions of atmospheric stability obtained by the layer method are identical with those obtained earlier by the particle method.

For different values of the vertical temperature gradient γ there are different ratios between the masses of saturated and unsaturated air drawn into circulation. For a small departure of the vertical temperature gradient γ from the moist adiabatic gradient γ_{wa} (the difference $\gamma - \gamma_{wa}$ is small) the ratio of the masses of moist and dry air which take part in the circulation should be small, i.e., the smaller the ratio $\left(\frac{\gamma - \gamma_{wa}}{\gamma_a - \gamma}\right)$, the smaller the mass of saturated air which can take part in the circulation.

Further analysis shows that in atmospheric layers containing both saturated and unsaturated air the development of convection is impossible (they are stable) if the temperature gradient γ of the layer is smaller than the moist adiabatic gradient. The appearance of convection is possible only for $\gamma > \gamma_{wa}$, in which case convection increases with increasing γ. However, we note that in the presence of clouds γ cannot exceed the value of the dry adiabatic gradient since it is prevented from doing so by the release of the latent heat of condensation. At the same time the active part in the development of convection belongs to moist air. Dry air plays a passive role and the development of active convection inside it is possible only in the absence of saturated air and for $\gamma > \gamma_a$.

From the preceding argument it is evident that in order to apply the layer method one needs to know the ratio between the descending and ascending masses of air, a difficult prerequisite which cannot always be fulfilled. An approximate determination of this ratio can be arrived at by estimating the size of clouds and open gaps between them. In those cases where it is applicable this method gives good results and is often used for studying questions relating to the development of convective clouds.

Part Two
RADIANT ENERGY

Chapter 7

RADIANT ENERGY IN THE ATMOSPHERE

§ 1. The fluxes of radiant energy in the atmosphere

In almost all the processes that take place in the atmosphere the sun is the principal source of energy. Of the enormous amount of energy continuously released by the sun into outer space, chiefly in the form of radiation, only one part in two billion ($\frac{1}{2.25 \cdot 10^9}$) penetrates the earth. This, however, is estimated to amount to $1.8 \cdot 10^{24}$ erg/sec ($\approx 2.5 \cdot 10^{18}$ cal/min). All other sources of terrestrial energy, external (stellar radiation, cosmic rays, etc.) as well as internal (heat from the earth's interior, radiation from radioactive substances, etc.), can be disregarded in comparison with the sun.

As it travels from the outer limits of the atmosphere to the earth's surface the radiant energy from the sun, or, as it is called, the **solar radiation**, experiences a series of important changes resulting from absorption and scattering. A considerable fraction, the so-called **direct solar radiation**, reaches the surface as a beam of parallel rays coming from the sun. At the same time, a certain fraction of the radiation scattered in the atmosphere reaches the surface as **scattered radiation** from all points in the sky. Taken together direct and scattered solar radiation constitute the so-called **total solar radiation**.

Part of the solar radiation reaching the earth's surface is reflected by this surface, but most of it is absorbed and used in heating the surface. We note here that while absorption of solar radiation does occur in the atmosphere it is much less significant there than at the earth's surface, since the atmosphere is principally a scattering medium. For example, of the total radiation absorbed by the earth as a whole the atmosphere absorbs only one quarter, the remaining three-quarters being absorbed by the surface. The fraction of solar radiation reflected by the surface and by the atmosphere (chiefly clouds) is called the **reflected radiation**.

On the other hand, the heated surface of the earth itself emits thermal radiation into the atmosphere. This is the so-called **earth emission**, or **earth radiation**. In turn, any part of the atmosphere also emits radiation. This **atmospheric emission** goes off in all directions, part of it reaching the surface and forming the **counter-emission of the atmosphere**. Some of it travels toward outer space and constitutes the **escaping emission of the atmosphere**. Earth radiation and atmospheric radiation, like solar radiation, are absorbed and reflected.

All these various streams of radiant energy have different spectral compositions Due to the high temperature of its emitter, the sun, the largest fraction of solar radiation is in the region of wavelengths shorter than

3–4μ, which includes the visible part of the spectrum (0.4–0.75μ). By contrast, the emission of the earth's surface and atmosphere is localized in the region of much longer wavelengths — practically, for actual conditions, in the region of wavelengths greater than 2μ. In view of this contrast in spectral composition, direct and scattered solar radiation are called short-wave radiation, and earth and atmospheric radiation long-wave radiation.

Thus in the atmosphere we find an entire system of fluxes of radiant energy differing in spectral composition and direction. In studying these fluxes we are concerned primarily with the following question: what amount of energy does each flux carry, and of this what fraction transforms into heat as a result of absorption? For any surface, the income and expenditure of radiant energy or, as it is commonly but not altogether felicitously called, the radiation balance, can be found by estimating the algebraic sum, from the energy standpoint, of all the fluxes reaching it.

The flux of radiant energy in the atmosphere, as we saw in Chapter 1, is studied in actinometry. The basic information about this branch of meteorology has been given in the present section.

Solar radiation, which includes the visible part of the spectrum, is simultaneously a source of light. Its study from this aspect, together with the study of the various optical phenomena which take place as it travels through the atmosphere, are grouped for convenience in a separate branch called atmospheric optics (see Part Six), which is the direct development and continuation of the present section.

§ 2. Fundamental quantitative characteristics of the radiation field

Radiant energy, like any other form of energy, may be expressed quantitatively in any energy unit. In meteorology it is customarily expressed in thermal units, namely calories (1 cal = $4.1855 \cdot 10^7$ erg = 4.18 watt/sec).

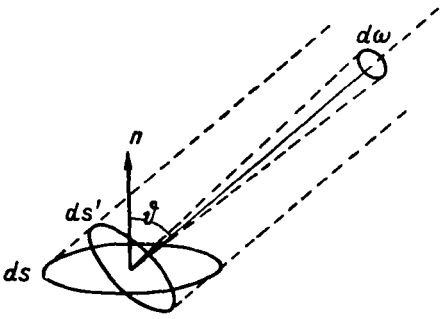


FIGURE 34

The fundamental quantitative characteristic of the radiation field is the radiation flux F, which we define as the amount of energy transported in all directions across a unit surface per unit time. If by $d\Phi$ we denote the

amount of energy in the wavelength interval between λ and $\lambda+\Delta\lambda$ transported across the element of surface ds in time dt, then, in conformity with the above definition, we can write the following expression for a monochromatic radiation flux:

$$F_\lambda = \frac{d\Phi}{ds\,dt},$$

and for the total (integral) radiation flux of all wavelengths

$$F = \int_0^\infty F_\lambda\, d\lambda.$$

In meteorology it is customary to measure the flux in cal/cm^2 · min; 1 cal/cm^2 · min = 0.6976 · 10^6 erg/cm^2 · sec = 0.0697 watt/cm^2.

In many cases one is interested in the flux dF_λ in the wavelength interval from λ to $\lambda+d\lambda$ across a surface element ds in time dt within the solid angle $d\omega$ in the direction r, making an angle ϑ with the outward normal n to the surface ds (Figure 34). Obviously,

$$dF_\lambda = I_\lambda \cos\vartheta\, d\omega\, dt\, d\lambda\, ds. \qquad (1)$$

The proportionality factor I_λ is the intensity of monochromatic radiation. From the relation (1) it is evident that I_λ is the quantity of radiant energy contained in a unit wavelength interval (from λ to $\lambda+1$) and unit solid angle ($d\omega=1$) crossing per unit time ($dt=1$) through a unit surface ($ds=1$) located at right angles to the direction of the beam of rays ($\vartheta=0°$).

If the intensity at a certain point does not depend on direction, the radiation field is called isotropic. If, moreover, the intensity remains the same

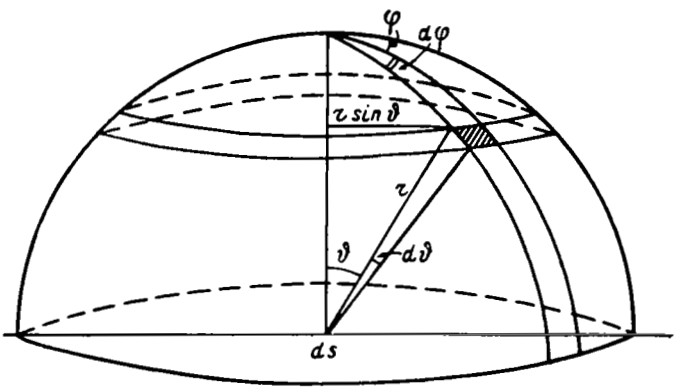


FIGURE 35

at all points of the field and in all directions, the radiation field is then called uniform and isotropic. The total, or integral, intensity of radiation I is obtained by integrating I_λ over all wavelengths, i.e.,

$$I = \int_0^\infty I_\lambda\, d\lambda. \qquad (2)$$

Intensity is the second most important quantitative characteristic of the radiation field. It is analogous to the quantity called luminance in photometry. In meteorology the intensity of radiation is expressed in $cal/cm^2 \cdot min \cdot sterad$.

Between the flux and intensity there exists a definite relationship which we will now obtain. To do this, we introduce the spherical coordinates r, φ, ϑ (Figure 35), and we write

$$d\omega = \frac{r \sin \vartheta d\varphi r d\vartheta}{r^2} = \sin \vartheta d\vartheta d\varphi$$

(where φ is the azimuth); then instead of (1) we obtain

$$dF_\lambda = I_\lambda \cos \vartheta \sin \vartheta \, d\vartheta \, d\varphi \, dt ds d\lambda. \tag{3}$$

The magnitude of the flux F_λ across the unit surface element ds is obtained by integrating the above expression over the hemisphere situated above the element ds (taking $ds = 1$, $dt = 1$)

$$F_\lambda = \int_0^{\frac{\pi}{2}} d\varphi \int_0^{2\pi} I_\lambda \cos \vartheta \sin \vartheta \, d\vartheta. \tag{4}$$

If the field is isotropic (I_λ does not depend on direction), then

$$F_\lambda = I_\lambda \int_0^{2\pi} d\varphi \int_0^{\frac{\pi}{2}} \cos \vartheta \sin \vartheta \, d\vartheta = 2\pi I_\lambda \frac{1}{2} = \pi I_\lambda, \tag{5}$$

while for the integral flux

$$F = \int_0^{\infty} F_\lambda d\lambda = \pi I, \tag{6}$$

i.e., for an isotropic radiation field, the flux from a hemisphere across an arbitrarily oriented surface is π times the intensity.

Next we define certain quantities characterizing the propagation of radiant energy in a medium. Let us take a mass element dm of volume dv in the medium which emits the same amount of radiant energy in all directions. Then the amount of energy dE_λ emitted by the mass element per unit time within the solid angle $d\omega$ and in the wavelength interval λ to $\lambda + d\lambda$ can be written as

$$dE_\lambda = e_\lambda \, dm \, d\omega \, d\lambda. \tag{7}$$

Here the coefficient e_λ is called the mass emission coefficient. Obviously, it is numerically equal to the amount of radiant energy emitted by a unit mass of substance per unit time within a unit solid angle and in a unit wavelength interval. For the total (integral) emission in all wavelengths we have the integral mass emission coefficient

$$e = \int_0^{\infty} e_\lambda d\lambda. \tag{7'}$$

As the energy emitted by a body passes through its bounding surface one often speaks of the emittance power of surfaces, always remembering, however, that it is the bodies and not the surfaces which radiate. Obviously, the definition of the emittance power of a surface is entirely analogous to

the definition of the intensity of radiation. Another concept which is frequently employed is that of the relative emissivity power. The latter is a dimensionless quantity representing the ratio of the emissive power of a given body to that of a black body at the same temperature.

Let us now consider the attenuation (extinction) of radiation intensity due to absorption. If a narrow beam of intensity I_λ moves along the path dl in a medium of density ρ, the attentuation of intensity in this process can be written as

$$dI_\lambda = -a_\lambda I_\lambda dl = -k_\lambda \rho I_\lambda dl, \tag{8}$$

where the quantity a_λ is the volume absorption coefficient and k_λ the mass absorption coefficient. In actinometry k_λ is usually expressed in cm^2/g and a_λ in cm^{-1}.

The absorption of radiant energy by surfaces (or rather by thin boundary layers) is determined by their absorptance. The latter is a dimensionless quantity numerically equal to the ratio of the absorbed radiation to the radiation incident on the given surface.

When radiation falls on a body it is partially reflected. This property of surfaces is characterized by the reflectance R_λ. The latter is a dimensionless quantity equal to the ratio of the intensity of radiation reflected by the surface to the intensity of radiation incident upon this surface. If the reflection is diffuse one speaks of the albedo of the surface; this is defined as the ratio of the reflected flux to the incident flux.

We stress here that both the absorptance and the reflectance of a body depend on wavelength. This property of absorbing and reflecting radiation only in definite wavelengths is called absorption (or reflection) selectivity. It is particularly conspicuous in the case of gases, the absorption spectra of which are generally highly complex. We note that for a black body the absorptance is unity.

For describing the absorption of radiant energy in its passage through a layer of matter one often introduces the so-called absorption functions, which give the integral properties of absorbing and emitting substances. In the case of directed radiation the absorption function is defined by the relation

$$A_I(m) = \frac{I(0) - I(m)}{I(0)} \tag{9}$$

and in the case of diffuse radiation analogously by

$$A_F(m) = \frac{F(0) - F(m)}{F(0)}. \tag{10}$$

Here m is the amount of absorbing substance in the layer while $I(0)$, $F(0)$ and $I(m)$, $F(m)$ are respectively the intensity and flux of the incident and emergent radiation.

Together with the absorption function the transmission function is also used. This is defined as

$$P(m) = 1 - A(m). \tag{11}$$

§ 3. Fundamental laws of emission

At the base of the emission of real bodies in nature lies Kirchhoff's law which relates the emittance and absorptance powers (and also the mass emission and mass absorption coefficients e_λ and k_λ) to each other. The Kirchhoff law states that, in thermodynamic equilibrium, the ratio of the mass emission coefficient of a certain substance to its mass absoprtion coefficient is a universal function of temperature and wavelength (for integral emission, only of temperature) and does not depend on individual properties of the substance (i.e., is the same for all bodies):

$$e_\lambda : k_\lambda = I(\lambda, T). \tag{12}$$

Since for black bodies $k_\lambda = 1$, it is obvious that the universal function $I(\lambda, T) = I_{\lambda, T}$ represents the intensity of black-body radiation for the same temperature and wavelength. With reference to processes of emission and absorption on surfaces it could also be stated that in thermodynamic equilibrium the ratio of the emittance of a surface to its absorptance depends only on temperature and wavelength and is equal to the intensity of black-body radiation.

Let us consider the fact that Kirchhoff's law was derived and formulated for the state of thermodynamic equilibrium. The latter is the state which any physical system will ultimately reach if it is maintained under unvarying external conditions and is left to itself. In this state every body in an enclosed system will emit as much energy as it absorbs. Consequently, for this state all bodies throughout the space should be at the same temperature.

In the atmosphere, where an open radiation field prevails, emission has a nonequilibrium character and thus Kirchhoff's law might seem to be inapplicable. However, it has been demonstrated that Kirchhoff's law does remain valid here provided that the velocity distribution be Maxwellian (a gas-kinetic temperature be definable) and that the energy levels responsible for absorption and emission be in thermal equilibrium according to Boltzmann's law. When applying Kirchhoff's law to the atmosphere we exploit the idea of the so-called local thermodynamic equilibrium, which is that state of the medium in which absorption and emission by every element of the medium at the temperature T takes place as if it were in a state of thermodynamic equilibrium at this temperature. In other words, we are assuming that the atmosphere is in a state of local thermodynamic equilibrium if at every point of the atmosphere one can define the local temperature in such a way that the emission and absorption coefficients e_λ and k_λ satisfy the relation $e_\lambda : k_\lambda = I(\lambda, T)$.

As special investigations have show, one may assume that the conditions of local thermodynamic equilibrium are fulfilled with sufficient accuracy to considerable altitudes (roughly to 50 km or even slightly higher) in the atmosphere.

The analytic form of the universal function $I_{\lambda, T}$ representing the emissive power of black bodies was given by Planck on the basis of ideas concerning the quantum character of the emission process. Planck's formula giving the intensity of black-body radiation is the following:

$$I_{\lambda, T} = \frac{2\pi c^2 h}{\lambda^5}\left(e^{\frac{ch}{k\lambda T}} - 1\right)^{-1} = \frac{c_1}{\lambda^5}\left(e^{\frac{c_2}{\lambda T}} - 1\right)^{-1}, \tag{13}$$

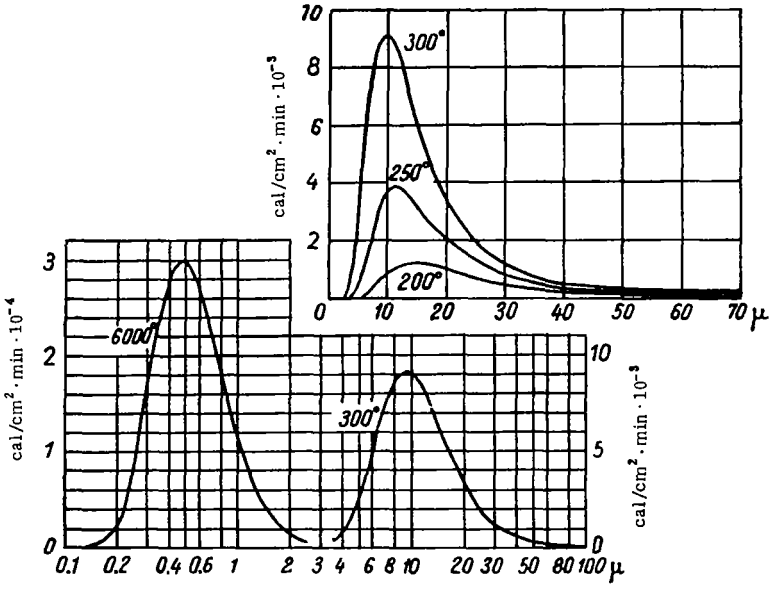


FIGURE 36. Distribution of energy in the spectrum of black-body radiation

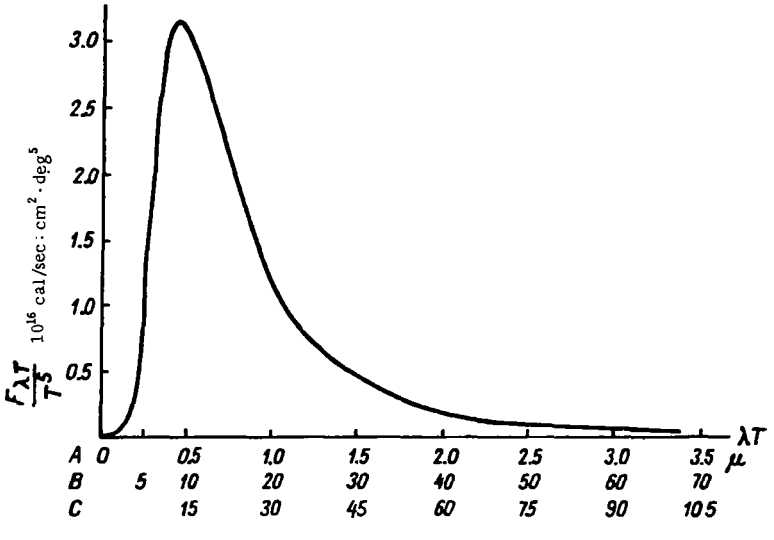


FIGURE 37. Black-body radiation

Scale A corresponds to T = 6000°, scale B to T = 300°, scale C to T = 200°.

where c is the velocity of light, $h = 6.62 \cdot 10^{-27}$ · erg/sec is Planck's constant and $k = 1.38 \cdot 10^{-16}$ erg/deg is Boltzmann's constant. The values of the constants c_1 and c_2 were found experimentally: $c_1 = 3.74 \cdot 10^{-5}$ erg/cm^2 · sec and $c_2 = 1.439$ cm · deg.

The distribution of the intensity of radiation over wavelengths for emitters at different temperatures is represented graphically in Figure 36. Writing Planck's formula (13) as

$$\frac{I_{\lambda, T}}{T^5} = \frac{c_1}{(\lambda T)^5}\left(e^{\frac{c_2}{\lambda T}} - 1\right)^{-1}$$

and plotting the values of λT and $\frac{\pi I_{\lambda, T}}{T^5}$ (in 10^{-16} cal/cm^2 · sec · deg^5 μ) along the abscissa and ordinate respectively, we obtain the curve of distribution of intensity given in Figure 37.

From these figures one can see that the intensity of emission increases very rapidly with rising temperature. Moreover, as the temperature increases the peak of the emission is shifted toward the shorter wavelengths.

The Wien displacement law gives the relation between the temperature of the emitting black body and the wavelength λ_{max} at which the intensity of radiation is maximum. It is expressed in the form

$$\lambda_{max} T = \text{const} = 2897\,\mu \cdot \text{deg} \qquad (14)$$

and indicates that the wavelength λ_{max} corresponding to the maximum intensity of emission $I_{\lambda, T}$ in the normal spectrum is inversely proportional to the absolute temperature of the emitter.

The Stefan-Boltzmann law. Since the flux of radiation is related to the intensity as $F_{\lambda, T} = \pi I_{\lambda, T}$, for the integral flux of black-body radiation, using (13), we find

$$F_T = \int_0^\infty F_{\lambda, T}\, d\lambda = \frac{2\pi^5 k^4}{15 c^2 h^3} T^4 = \sigma T^4, \qquad (15)$$

where σ is the stefan-Boltzmann constant ($\sigma = 5.669 \cdot 10^{-12}$ watt/cm^2 · deg^4 = $= 0.814 \cdot 10^{-10}$ cal/cm^2 · min · deg^4).

Let us draw a few conclusions from the above regarding the problems under consideration:

1) in black-body emission at the temperature $T = 6000°$ (emission temperature of the sun) about 99% of the emitted energy lies in the region of the spectrum between 0.17 and 4.0μ ($\lambda_{max} = 0.475\mu$);

2) at $T = 288°$ (mean temperature of the earth's surface) the emission spectrum is limited to the 3–80μ interval ($\lambda_{max} = 10\mu$);

3) at temperatures of the order of 210–220° (temperature of the stratosphere) the spectrum is confined to wavelengths of 4–120μ ($\lambda_{max} \approx 14$–15μ).

§ 4. The sun and its radiation. The solar constant

The study of the sun and its radiation is one of the principal tasks of astrophysics. At the same time the question of the amount of solar energy received by the earth is of very great importance for meteorology as well.

The astronomical data on the sun which we require are given very briefly in the following.

The sun, the star closest to us, is a yellow dwarf. Its linear dimensions can be pictured from its mean radius, which is 695,500 km — 109.1 times the radius of the earth. The distance of the earth from the sun changes over the year owing to the ellipticity of the earth's orbit. The earth is nearest the sun (distance $147 \cdot 10^6$ km) on January 2 at the point of the orbit called the perihelion, and farthest ($152 \cdot 10^6$ km) on July 5 at the point of the orbit called the aphelion. The mean distance of the earth from the sun (length of the semi-major axis of the earth's orbit) is $149.5 \cdot 10^6$ km. Despite the enormous linear dimensions of the sun the apparent size of the solar disk is small owing to its great distance; its mean apparent angular diameter is 31'59" (32'31" at perihelion and 31'27" at aphelion).

The sun's structure is highly complex. Its interior regions which are inaccessible to direct observation have a very high temperature, estimated to be many tens of millions of degrees (up to $20 \cdot 10^6$ deg). They are the site of the complex nuclear reactions that are the source of solar energy. The outer layer of the sun, which in practice emits nearly all the observed solar radiation, is called the photosphere. Its thickness does not exceed 0.001 of the solar radius and amounts to about 200–300 km. The temperature of the surface of the photosphere is about 6000°. Above the photosphere lies the luminous but nearly transparent atmosphere of the sun, the so-called chromosphere, which extends to a height of the order of 15,000 km. Above the chromosphere to very large distances (millions of kilometers) extends the outermost envelope of the solar atmosphere — the so-called corona.

The gases which constitute the sun are in a state of continual violent motion, and each of the sun's envelopes is the site of constantly developing processes; this is manifested in the fact that the observed surface of the sun is not uniform. Thus the surface of the photosphere appears to consist of a large number of individual bright granules separated by dark spaces and describing continuous and irregular motions. The granulation is due to the presence of rising and descending fluxes in the photosphere. Dark formations, the so-called sunspots, are seen in some parts of the surface. Also observed here, near the edge of the solar disk, are bright fibrous formations called faculae. Highly variable formations known as flocculi, chromospheric flares, prominences, etc., are found in the chromosphere and corona. The set of all such phenomena, which are closely interrelated, constitutes the highly variable solar activity. Prolonged observation has established the presence of a definite cyclicity in the development of solar activity. The eleven-year cycle, the existence of which has long been known, is the one most thoroughly studied.

Upon changes in solar activity modifications are also observed in the intensity of a number of geophysical phenomena and especially in those controlled by processes taking place in the upper atmospheric layer (ionization of the ionosphere, aurora polaris, magnetic storms, etc.). The question of the relationship between solar activity and processes in the atmosphere is a very important one and has great practical significance: much, however, remains to be investigated, especially regarding the connection between solar activity and meteorological processes developing in the lower atmosphere.

The various complex processes taking place in each of the sun's envelopes are accompanied by the emission of energy. On the principal emission of the photosphere, which corresponds fairly closely to black-body

emission at a temperature close to 6000°K, are superposed the emission of the chromosphere and corona and that of a few of the above-mentioned formations. As a result of all these processes of emission— and also absorption in the solar atmosphere— the composition of the solar radiation reaching the earth is highly complex. Investigation into the composition of solar radiation is seriously hampered by the fact that many of its components are weak and are absorbed in the upper layers of the earth's atmosphere, thus becoming inaccessible to observation from the ground. The use of rockets and special high-sensitivity apparatus (photon counters, special filters, radiophysical methods, and so forth) has lately furnished a new set of data which is steadily being amplified.

It has now been established by observation that in addition to its electromagnetic emissions, which cover a broad range of wavelengths (from 1 Å to several tens of meters), the sun also sends off a flux of corpuscular radiation consisting of electrically charged particles (mainly protons and electrons) moving at speeds of 400 to 3000 km/sec. However, the flux of corpuscular radiation is very small; on the average it is estimated to be of the order of 10^{-7} cal/cm^2 min. As it is absorbed in the atmosphere at heights of over 100 km, corpuscular radiation is of considerable significance for the development of many phenomena observed in the upper layers of the earth's atmosphere. Among the electromagnetic emissions of the sun, it has been established that there is also a component of soft X-rays in the region around 7 Å, detectable at heights over 50 km, as well as radio waves in the region from 0.8 to 20 cm; our atmosphere is transparent to the latter and allows them to reach the surface. The study of the radio emission of the sun and other celestial bodies is the subject of a new branch of astronomy, radio astronomy.

TABLE 19

Individual regions of the solar spectrum

Name	Wavelength interval (μ)	Magnitude of flux		Remarks
		10^{-3} cal/cm^2·min	%	
Ultraviolet region:	0.20–0.40	140	7	
C	0.20–0.28	8	0.4	Does not reach earth's surface
B	0.28–0.32	25	1.2	Also called the Dorno region
A	0.32–0.40	107	5.4	Reaches earth's surface
Visible region:	0.40–0.75	910	46	
A	0.40–0.52	350	18	Violet to green rays
B	0.52–0.62	300	15	Green to red rays
C	0.62–0.75	260	13	Red rays
Infrared region:	0.75–24	930	47	
A	0.75–1.4	640	32	Near infrared region
B	1.4–3.0	250	13	
C	3.0–24	40	2	Middle infrared region
Total	0.20–24	1980	100	

Rocket exploration has also established the presence in solar emission of several bands in the far-ultraviolet region of the spectrum with wavelengths

less than 1900 Å. This emission is particularly intensive in the region around 1000 Å.

So little energy is supplied in these regions, however, that they present no interest to anyone evaluating the total quantity of heat and light received by the earth from the sun.

In actinometric investigations we are concerned with the region of the continuous spectrum of solar radiation between 0.19 and 24μ, of which only those rays in the 0.29–24μ interval reach the earth's surface. Table 19 presents the conventional breakdown of this broad spectral region into narrower individual regions; certain characteristics of the latter are also given.

In the short-wave region ($\lambda<0{,}29\,\mu$), thanks to rocket studies, the interval between 0.22 and 0.29 has been investigated fairly thoroughly (quantitatively). As far as the region from 0.22 to 0.19μ is concerned only a few qualitative data are available.

In the infrared region, as we indicated, only radiation with a wavelength up to 24μ reaches the earth's surface. The earth's atmosphere is apparently nontransparent to radiation with wavelengths greater than 24μ, and so far these rays have been observed neither at the earth's surface nor in rocket studies.

Detailed investigation of the distribution of energy in the spectrum of solar radiation is hampered by the necessity of accounting for absorption in the atmosphere (for details see Chapter 8, § 3) and also by the fact that not all rays reach the earth's surface. Figure 38 shows the distribution of energy in the normal solar spectrum in the region between 0.29 and 2.2μ based on ground observations.

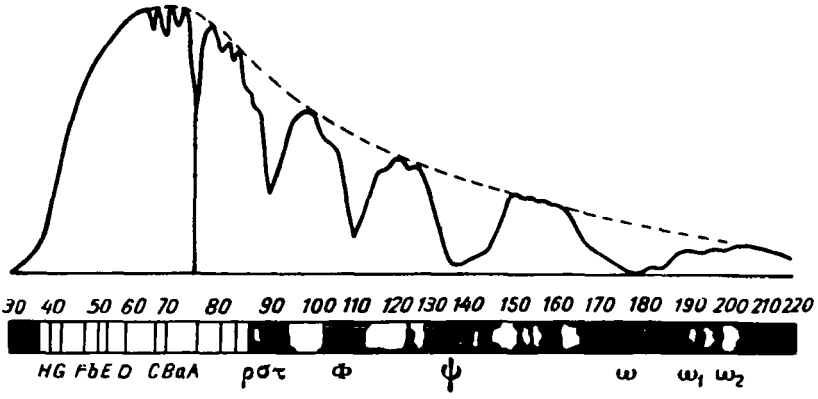


FIGURE 38. Distribution of energy in the normal solar spectrum

The spectrum observed at ground level is continuous but displays a vast number of absorption bands and lines; the principal ones are indicated on the figure. There are many thousands of such absorption (Fraunhofer) lines and bands. They are largely due to absorption of photospheric radiation in the solar atmosphere and only a certain proportion is caused by absorption in the earth's atmosphere (so-called telluric absorption lines). At the earth's surface the solar spectrum is sharply cut-off on the side of

the short wavelengths owing to strong absorption of rays with $\lambda < 0.29\,\mu$ by ozone and oxygen in yet higher atmospheric layers. The distribution of energy in this ultraviolet region of the spectrum (based on rocket data) is given in Figure 39.

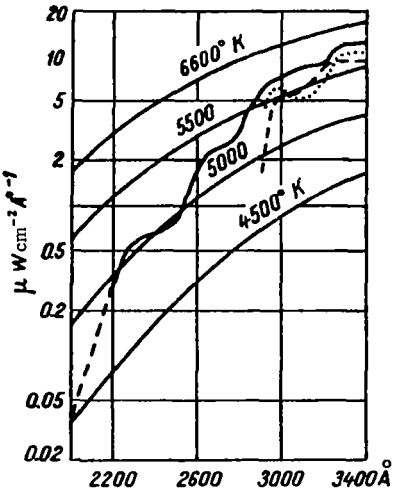


FIGURE 39. Ultraviolet spectrum of the sun from rocket data

Data obtained by various investigators have made it possible to construct the curve of energy distribution outside the atmosphere (Figure 40). For comparison the distribution of radiation in the black-body spectrum for a temperature of 6000°K is shown in the same figure by a broken line. One can see that the maximum energy in the solar spectrum corresponds to a wavelength of about $0.49\,\mu$ and that the solar distribution differs significantly (especially in the ultraviolet region) from the Planck distribution for a temperature of 6000°. Numerous attempts to fit it to the Planck distribution curve for some other temperature have shown that this cannot be done, though certain sections can be fitted fairly closely to the curves of emission at several different temperatures. Thus the best fit in the infrared region is with the curve for $T = 5779°K$, and in the visible region for $T = 6075°K$. All this shows that the sun's emission is not to be identified with black-body emission. Many authors give this distribution in tabular form. V. G. Kastrov expressed it by an empirical formula

$$e_{0,\lambda}\,\Delta\lambda = 0.021\,\lambda^{-2.3}\,e^{-0.0327\Delta\lambda}, \qquad (16)$$

expressing the flux in 10^{-3} cal/cm^2 min in the spectral interval $\Delta\lambda = 0.02\,\mu$.

If one knows the curve of distribution of the solar energy incident on a surface perpendicular to the sun's rays outside the atmosphere, one can determine the size of the incident flux. Indeed, the area bounded by the axis of abscissae (wavelength) and the curve of energy distribution in a definite wavelength sector from λ_1 to λ_2 gives, in relative units, the quantity of energy incident in this wavelength interval. The term solar constant (S_0) is applied by convention to the total amount of radiant energy transmitted per unit time (1 min) across a unit surface (1 cm^2) situated at right angles to the sun's rays and located at the mean distance from the sun.

The solar constant can also be defined in a slightly different way, as the magnitude of the flux of radiant energy from the sun onto a surface situated at right angles to the rays outside the atmosphere at the earth's mean distance from the sun.

Since the earth's distance from the sun, as we mentioned earlier, changes over the year, the flux of solar radiation S' outside the atmosphere at a certain instant for which the earth's distance from the sun is R will be different from S_0 and is related to S_0 as $S_0 = \frac{R^2}{R_0^2}\,S'$, where R_0 is the mean distance of the earth from the sun. The variation of the solar constant

through the year amounts to ± 3.4% of its mean value. It is maximum when the earth is at perihelion (beginning of January) and minimum at aphelion (beginning of July).

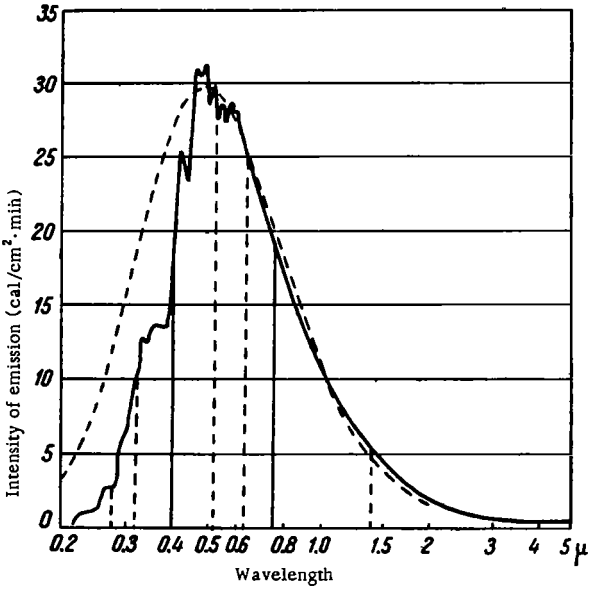


FIGURE 40. Distribution of energy in the solar spectrum outside the atmosphere

Dashed lines indicate black-body emission at T = 6000°K.

The determination of the exact absolute value of the solar constant is of first-rate importance and a large number of researches has been devoted to this question. Without dwelling on the specialized question of the methods used for the experimental determination of the solar constant, we note that it now appears from work carried out in the last decade that the value previously accepted on the basis of work by American astrophysicists, $S_0 = 1.94$ cal/cm^2 · min, is too low; the more probable value is about 2 cal/cm^2 · min. The international commission which discussed this question in Toronto in 1957 suggested using the value $S_0 = 1.98$ cal/cm^2 · min for work during the International Geophysical Year.

Long series of systematic measurements provide an answer to the important question of the degree to which the solar constant remains constant in time and whether or not it experiences fluctuations related to changes in solar activity. Observations show that the measured values of S_0 display no systematic changes in time and that whatever small fluctuations are recorded lie within the accuracy of measurement. At the same time reasons of principle preclude our expecting large variations in the values of S_0, since this would presuppose the observation of very large variations in the sun temperature. However, the short-lived flares occasionally observed in the ultraviolet region of the solar spectrum (especially during years of increased solar activity) are too small and too brief from the energy standpoint to produce changes in S_0 detectable with the present accuracy of observation.

Chapter 8

INFLUENCE OF THE ATMOSPHERE ON FLUXES OF RADIANT ENERGY

§ 1. Fundamental relations. The path of a solar ray in the atmosphere

When fluxes of radiant energy are propagated through the atmosphere they are weakened by absorption and scattering. Scattering predominates in the visible region of the spectrum but outside the visible spectrum, both in the ultraviolet and in the infrared region, the atmosphere is chiefly an absorbing medium. Let us consider the total attenuation of the flux of radiant energy in the atmosphere. To do this we will take a directed beam of monochromatic rays incident at right angles on an atmospheric layer of thickness l.

If we denote the flux by F_λ, its attenuation on travelling over the path dl in the atmosphere will be given by

$$dF_\lambda = -\alpha_\lambda F_\lambda dl, \tag{1}$$

where the factor α_λ is called the coefficient of attenuation (extinction coefficient); its dimensionality is cm^{-1}. If the active factor which produces the attenuation is scattering alone α_λ is called the coefficient of scattering; for the case of absorption only α_λ is the coefficient of absorption. It depends on a number of factors but in the first approximation one can consider it as being proportional to the density of the atmosphere or, correspondingly, to the concentration of the absorbing substance ρ, i.e., one can write

$$\alpha_\lambda = k_\lambda \rho,$$

where k_λ is the mass attenuation coefficient. Its dimensionality is $cm^2\ g^{-1}$.

Then instead of (1) we obtain

$$dF_\lambda = -k_\lambda \rho F_\lambda dl. \tag{2}$$

If the ray travels over a finite path l then, integrating (1), we obtain

$$F_\lambda = F_{\lambda,0}\, e^{-\int_0^l k_\lambda \rho dl} = F_{\lambda,0}\, e^{-\int_0^l \alpha_\lambda dl}, \tag{3}$$

where $F_{\lambda,0}$ and F_λ are the fluxes at the beginning and end of the path.

The quantity

$$\Theta = \int_0^l \alpha_\lambda dl = \int_0^l k_\lambda \rho dl \tag{4}$$

is usually called the optical mass or thickness of the layer, and the quantity $\alpha_\lambda = k_\lambda \rho$ the optical density. It is easy to show that a layer with the optical thickness $\Theta = 1$ will attenuate the flux by a factor of e or by $\left(1 - \frac{1}{e}\right) 100\%$ of its initial value.

Cases where the atmosphere is optically homogeneous along the path of the rays, i.e., $\alpha_\lambda = k_\lambda \rho = \text{const}$, are encountered only exceptionally. In particular, this may be assumed true if the rays are propagated horizontally and their path is fairly short. Usually, and especially for rays propagated at a slant in the atmosphere, the problem of determining the attenuation of the flux is distinctly more complex and reduces to calculating the integral $\int_0^l \alpha_\lambda \, dl$ for variable α_λ.

In this connection let us consider the question of the path of a solar ray in the atmosphere.

Turning to formula (3), we introduce the symbol S_λ for the monochromatic flux of direct solar radiation, i.e., we write equation (3) in the form

$$S_\lambda = S_{\lambda,0} e^{-\int_0^\infty \alpha_\lambda \, dl} = S_{\lambda,0} e^{-\int_0^\infty k_\lambda \rho \, dl}. \qquad (5)$$

Here S_λ and $S_{\lambda,0}$ are respectively the values of the flux of direct solar radiation upon a surface perpendicular to the beam at the earth's surface and at the upper boundary of the atmosphere.

According to (5), the quantity $\int_0^\infty k_\lambda \rho dl = \Theta$ is the optical thickness of the atmosphere. It is obviously dependent on the zenith distance of the sun $z_\odot$, and the problem reduces to determining $\Theta(z_\odot)$ along the path of the ray. We will proceed to do this in the following way.

We determine the optical mass for the case of vertical incidence ($z_\odot = 0$, $l = h$), i.e., we determine the quantity $\Theta(0)$. It is obvious that

$$\Theta(0) = \int_0^\infty k_\lambda \rho dh. \qquad (6)$$

Under the condition that k_λ and ρ be independent of h, we obtain for $\Theta(0)$ the simple formula $\Theta(0) = k_\lambda \rho_0 H$, where ρ_0 and H are the density and height of the homogeneous atmosphere (scale height).

We now set up the ratio

$$m(z_\odot) = \frac{\Theta(z_\odot)}{\Theta(0)} = f(z_\odot). \qquad (7)$$

If $\Theta(0)$ is known then, knowing the values of $m(z_\odot) = f(z_\odot)$, one can determine $\Theta(z_\odot)$ for any $z_\odot$. The function $m(z_\odot)$ is called the atmospheric "mass" (relative air mass). It should be emphasized that this concept of mass has nothing in common with usual ideas concerning mass and that the quantity $m(z_\odot)$ is dimensionless. Taking the optical thickness of the atmosphere in the vertical direction as unity, i.e., setting $\Theta(0) = 1$, one can find the value of $\Theta(z_\odot)$ in relative units for any zenith distance $z_\odot$.

We will now show how $m(z_\odot)$ is calculated. The simplest way of solving the problem is to disregard the curvature of the earth's surface and the distortion of the rays' path due to refraction in the atmosphere (see Chapter 26). It is easy to see that the mass $m(z_\odot)$, as the ratio of the optical

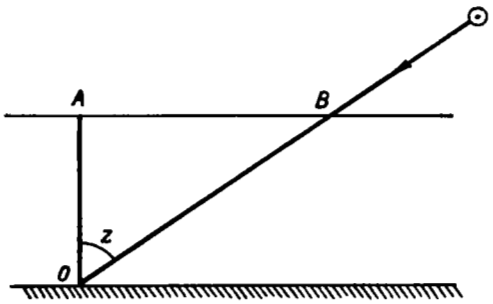


FIGURE 41

mass, is at the same time the ratio of the path lengths of a ray in an inclined direction (l) and in the vertical direction (h). In this case (Figure 41) it is obvious that

$$m(z_\odot) = \frac{dl}{dh} = \sec z_\odot. \qquad (8)$$

It appears that formula (8) can be used with sufficient accuracy to values of $z_\odot = 60°$. For a more exact determination of $z_\odot$ (for $z_\odot > 60°$) one must allow for the curvature of the path of the rays and also for the curvature of the atmosphere. This was done by a number of authors (Laplace, Bemporad, Staudt, et al.) and has led to more exact expressions for the function $m(z_\odot)$. The data obtained are compared in Table 20.

TABLE 20

Values of the atmospheric mass $m(z_\odot)$

$z_\odot$		0	30	60	75	85	87	88	90
$\sec z_\odot$		1.00	1.15	2.00	3.86	11.48	19.11	28.65	∞
$m(z_\odot)$	after Laplace	1.00	1.15	1.99	3.81	10.20	14.84	18.84	
	after Bemporad	1.00	1.15	2.00	3.82	10.40	15.36	19.79	$\sim$ 35-40

From these data one can see that for small zenith distances the values of $m(z_\odot)$ change slowly with changing $z_\odot$; strong variation of $m(z_\odot)$ is observed only for large values of $z_\odot$. For small changes of $z_\odot$ the accuracy of determination of $m(z_\odot)$ for $z_\odot > 85°$ does not exceed 0.1 [sic]; it is particularly small for $z_\odot$ close to 90°.

The estimates of atmospheric mass presented here apply to the case of observation at sea level, where one can take the pressure $T = 760$ mm and temperature $p \approx 273°$. However, the pressure at the earth's surface often deviates significantly from 760 mm Hg, and observations are sometimes

carried out at a certain height above sea level. In this case one must obviously correct the tabulated values of $m(z_\odot)$ by multiplying them by $\frac{p}{760}$, where p is the pressure at the point of observation, i.e.,

$$m'(z_\odot) = \frac{p}{760} m(z_\odot). \tag{9}$$

This is absolutely necessary in the case of observations at a considerable height above sea level. We note for example that at a height of 5.5 km, where the pressure is of the order of $p = 380$ mm Hg, the mass of the atmosphere for $z_\odot = 0$ is 0.5.

§ 2. Integral attentuation of the flux of direct solar radiation in the atmosphere

From the preceding section we have the following expression for the optical mass of the atmosphere along the path of a solar ray for the zenith distance of the sun $z_\odot$:

$$\theta(z_\odot) = \theta(0) m(z_\odot) = \int_0^\infty k_\lambda \rho dl. \tag{10}$$

The relation (5) can thus be written as

$$S_\lambda = S_{\lambda,0} e^{-\theta(0) m(z_\odot)}. \tag{11}$$

We set

$$e^{-\theta(0)} = p_\lambda. \tag{12}$$

The quantity p_λ is called the transmission coefficient of the atmosphere for radiation of wavelength λ, or, in other words, the monochromatic transmission coefficient. Let us assume that k_λ does not depend on the path l. Then

$$p_\lambda = e^{-k_\lambda \rho_0 H} = e^{-a_\lambda H}. \tag{13}$$

Writing (11) in terms of p_λ we obtain

$$S_\lambda = S_{\lambda,0} p_\lambda^m. \tag{14}$$

This formula holds only for monochromatic beams. Taking its logarithm we obtain

$$\ln S_\lambda = \ln S_{\lambda,0} + m \ln p_\lambda, \tag{15}$$

i.e., there exists a linear relationship between $\ln S_\lambda$ and m.

In equation (15) the quantities S_λ and m are determined from observations. The quantities $S_{\lambda,0}$ and p_λ are unknown but can be determined from two observations for different values of the mass traversed by the rays (m_1 and m_2), assuming that the optical properties of the atmosphere were the same in these observations. Indeed, from the two equalities

$$S_{\lambda,1} = S_{\lambda,0} p_\lambda^{m_1} \quad \text{and} \quad S_{\lambda,2} = S_{\lambda,0} p_\lambda^{m_2}$$

we easily obtain

$$S_{\lambda,0}=S_{\lambda,1}\left(\frac{S_{\lambda,1}}{S_{\lambda,2}}\right)^{\frac{m_1}{m_2-m_1}} \tag{16}$$

and

$$p_\lambda=\sqrt[m_1-m_2]{\frac{S_{\lambda,1}}{S_{\lambda,2}}}, \tag{17}$$

i.e., we can determine both $S_{\lambda,0}$ and p_λ.

Usually, however, the direct solar radiation is measured either over its entire wavelength range or in a certain more or less extensive region of the spectrum bounded by the wavelengths λ_1 and λ_2. In the first case one should clearly write

$$S_m=\int_0^\infty S_\lambda\,d\lambda=\int_0^\infty S_{\lambda,0}\,p_\lambda^m\,d\lambda, \tag{18}$$

where S_m is the total flux of direct solar radiation at the level of the earth's surface in the case where the mass of the atmosphere in the direction of the sun is m.

In practical work a relation analogous to (14) is frequently used for the total flux in place of the exact relation (18):

$$S_m=S_0p_m^m, \tag{19}$$

where p_m is a certain averaged (generalized) integral transmission coefficient of the atmosphere for a mass of the atmosphere equal to m.

The relation (19) is also the one usually used to calculate p_m from data of actinometric observations; it follows from (19) that

$$p_m^m=\frac{S_m}{S_0}=\frac{\int_0^\infty S_{\lambda,0}p_\lambda^m\,d\lambda}{\int_0^\infty S_{\lambda,0}\,d\lambda}. \tag{20}$$

Even simple qualitative considerations point to the conclusion that the integral transmission coefficient p_m depends on the mass m crossed by the rays. Indeed, the monochromatic coefficient of transmission p_λ in the atmosphere is considerably smaller for short wavelengths than for long ones; therefore, if a beam of complex radiation passes through the atmosphere, the short waves inside it will be attenuated to a greater extent than the long waves. As a result when m increases the composition of the beam changes and its "optical center" is shifted toward the longer waves, i.e., m increases with increasing p_m.

Bearing in mind the dependence of p_m on the mass traversed, Kastrov proposed replacing formula (19) for S_m by the following formula which he derived by analyzing observational results:

$$S_m=\frac{S_0}{1+cm}, \tag{21}$$

where c is a certain coefficient the numerical value of which is assumed to be independent of m. However, observations show that the values of c also depend to some extent on m, though this dependence is slight, especially for small values of m.

§ 3. Absorption of radiant energy in the atmosphere

As we mentioned before, the attenuation of the flux of radiant energy in the atmosphere is a consequence of its absorption and scattering, absorption being dominant in the ultraviolet and infrared regions of the spectrum.

The principal absorbers of radiant energy in the atmosphere are oxygen, ozone, carbon dioxide and water vapor, and also dust.

Oxygen (O_2) has absorption bands in the visible and chiefly in the ultraviolet region of the spectrum. In the visible part of the spectrum the absorption bands are: band ***A***, with center around 0.76 μ, and band ***B***, with center around 0.69 μ. Absorption in these major bands is limited and therefore their influence on the attenuation of solar radiation is also small. Of far greater significance are the systems of absorption bands located in the ultraviolet regions of the spectrum, e. g., the Herzberg bands in the region 2600–2420 Å, which grade into the continuum extending to 2000 Å, and the Runge-Schumann bands in the 1925–1760 Å region with a very large absorption coefficient. Next come bands of intensive absorption extending to 1350 Å.

Absorption in the ultraviolet region of the spectrum is so strong that detailed investigation into feeble solar radiation has so far proved impossible even with the aid of observations at the great altitudes reached by rockets. Absorption of solar radiation by oxygen in this region of the spectrum ($\lambda < 2420$ Å) produces dissociation of molecular oxygen and the formation of ozone in upper atmospheric layers, as well as ionization of gases in still higher layers. Finally, absorption of solar radiation by oxygen may also affect the thermal regime of the upper atmospheric layers.

Ozone (O_3), which occurs in a layer extending from 10 to 60 km in height and is concentrated at about 22 km, has extremely high absorptivity in certain regions of the spectrum and especially in its ultraviolet region. The principal absorption bands of ozone in the ultraviolet and visible regions of the spectrum are illustrated schematically in Figure 42, which also shows the major absorption bands of oxygen mentioned above. The most important absorption band of ozone, the so-called Hartley band, lies in the ultraviolet region between $\lambda = 200$ mμ to $\lambda = 320$ mμ. Immediately next to it, almost in continuation, lies the weaker Huggins band which extends to $\lambda = 360$ mμ.

In the Hartley band maximum absorption occurs at the wavelength $\lambda = 255$ mμ, where the decimal absorption coefficient* has the value $a = 145\ \text{cm}^{-1}$. Absorption in this region is so strong that an ozone layer 0.25 μ thick (for $p = 760$ mm Hg and $t = 0°$) will attenuate radiation by one half.

In the visible part of the spectrum we find the Chappuis band system (430–750 mμ) with maximum absorption at about $\lambda = 600$ mμ, where the absorption coefficient amounts to $a = 0.068\ \text{cm}^{-1}$. Thus absorption in this region is weak. However, if it is recalled that this band occurs in the region having the maximum energy of the solar spectrum, one sees that in absolute units the attenuation of the flux of direct solar radiation due to this band is pronounced.

The first and most important consequence of the strong absorption in the ultraviolet by ozone is that the solar spectrum observed at the surface, as we noted before, is cut-off abruptly at a wavelength of about 300 mμ. On

* $a = \alpha \lg e = 0.4343\alpha$.

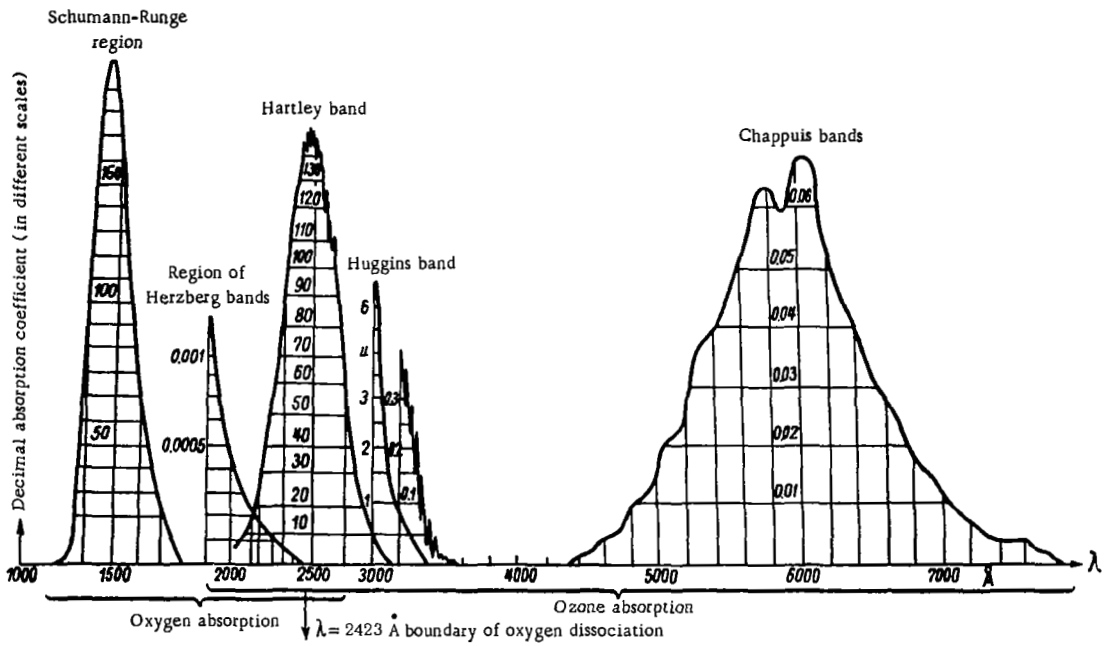


FIGURE 42. Absorption coefficients of ozone and oxygen in the ultraviolet and visible regions of the spectrum (in different scales)

the average one may assume that for high solar elevations and low ozone contents in the atmosphere the boundary of the solar spectrum occurs at about 291 mμ; the shortest wavelength recorded at the surface is 288.6 mμ. As one moves upward over the surface there is a certain prolongation of the solar spectrum toward the shorter waves, but this phenomenon is observed only from a height of 50–55 km.

Calculations of the absorption of solar radiation by ozone in these spectral regions show this to be about 2–3% (tentatively) of the entire integral flux.

Finally, ozone has several absorption bands in the infrared region of the spectrum. Under atmospheric conditions, however, most of these bands are overlapped by the intense absorption bands of carbon dioxide and water vapor, and only a narrow (9.4–9.9 μ) though intense band centered at $\lambda = 9.65\,\mu$ has any serious significance.

Carbon dioxide (CO_2) has a number of absorption bands in the infrared region of the spectrum. The strongest of these is a narrow band centered at $\lambda = 4.3\,\mu$. This band is of limited significance, however, as it is located at the extreme edge of the curve of solar radiation; moreover, earth emission is also limited here. A more important band is the broad one (12.9–17.1 μ) centered about $\lambda = 14.7\,\mu$; though the absorption at its maximum is weaker than in the band near $\lambda = 4.3\,\mu$, it is broad and is situated at the maximum thermal emission of the atmosphere. In addition to these two major bands carbon dioxide also has a series of weak bands, e.g., those with centers at about 2.7 and 2.05 μ, and also between 1 and 2 μ.

Water vapor (H_2O) is the most important factor in the absorption of radiant energy in the atmosphere, not only because it is present in large amounts but also because of the very great number of lines and bands in its complex absorption spectrum. A significant number of the absorption lines of water vapor are found in the visible part of the spectrum, but all are relatively weak; of these we mention only band *a* (730–685 mμ) and the so-called "rain"-band (606–585 mμ).

Of far greater significance are the bands located in the near and especially the far-infrared regions of the spectrum. These absorption bands, the most intensive ones in the spectrum of solar radiation, are listed in Table 21.

TABLE 21

Absorption bands of water vapor

Symbol	α	β	ρστ	Φ	ψ	Ω
Band center (μ)	0.72	0.82	0.93	1.13	1.38	1.86
Symbol	ω_1	ω_2	x	—	y	
Band center (μ)	2.01	2.05	2.68	3.2–4.0	4.0–4.9	

The magnitude of the absorption in all these bands (cf. Figure 43) depends on the water vapor content and increases with the latter. This being the case, the total absorption of solar radiation by water vapor naturally varies with its content. Among the many formulas suggested for the magnitude of absorption of the flux of solar radiation we cite the empirical formula of Mügge and Möller

$$\Delta S = 0.172\,(m w_0)^{0.303}, \qquad (22)$$

where ΔS is the absorption of the flux of solar radiation in cal/cm^2 · min, m is the mass of the atmosphere in the direction of the sun and w_0 the total water vapor content in a vertical column of air of unit section.

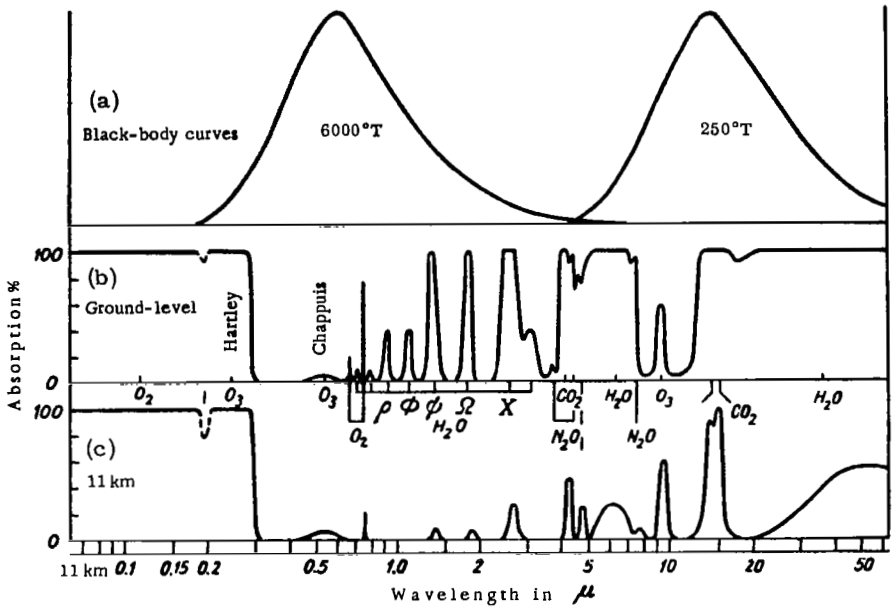


FIGURE 43. Absorption spectrum of the earth's atmosphere

The absorption spectrum of water vapor is particularly complex in the far-infrared region, a fact of tremendous significance for the study of long-wave radiation. A great many investigations, experimental as well as theoretical, have been carried out in connection with this question. A very important result of these studies is that, with the exception of the interval between roughly 8 and 12 μ, radiant energy is almost entirely absorbed by water vapor in the region from 4 to 40 μ. Absorption by water vapor begins again at $\lambda = 12\mu$ and increases strongly at wavelengths of the order of 15–20 μ. We recall that the broad carbon dioxide band mentioned earlier lies in the 12.9–17.1 μ region. The 8–12 μ wavelength interval is frequently called the atmosphere's transmission window, or simply window. The general features of the infrared absorption spectrum of water vapor are clearly seen in the middle section of Figure 43. The figure shows the total absorption spectrum of the atmosphere as a whole at the earth's surface (curve b) and at a height of 11 km (curve c). The upper curve (a) gives the curves of black-body emission for temperatures of 6000° T (solar radiation) and 250° T (thermal emission of the atmosphere) [drawn with equal areas].

From the figure (curve c) one sees that even at great altitudes the influence of water vapor remains considerable despite its low content.

The dust contained in the atmosphere also exercises a considerable influence on the absorption of radiant energy, as one can judge from the weakening of the flux of solar radiation experienced whenever the atmosphere is polluted by dust, smoke, etc. Another manifestation of this is seen in

industrial cities, where the flux of direct solar radiation is usually smaller (by several percents) than outside town limits. The importance of dust as an attenuator of solar radiation can also be verified during forest and peat fires and after strong volcanic eruptions.

Finally, the nitrogen oxides, sulfur compounds and many other substances present in the atmosphere have absorption bands in the long-wave region of the spectrum. However, although the absorption coefficients of some of these are considerable, their influence can be disregarded because they are contained in small amounts.

Taking the absorption of solar radiation as a whole, one can say that solar radiation is attenuated in the atmosphere by a factor of 17–25% on the average and that most of it (~75–80%) penetrates to the earth's surface.

§ 4. Scattering of radiant energy in the atmosphere

The second factor responsible for the attenuation of radiant energy in the atmosphere is scattering. It is well known that radiant energy undergoes scattering whenever the medium through which it propagates is optically inhomogeneous (in particular, when foreign particles with properties different from those of the surrounding medium are present, and also in the presence of density fluctuations in the medium).

The essence of the scattering process is as follows: under the influence of the oscillating electric vector of the incident electromagnetic wave, the electrons in the scattering particle begin to describe forced oscillations and emit waves. Consequently, the scattering particle itself becomes a source of electromagnetic waves. In general the set of electromagnetic waves emitted by the particle is highly complex. However, if the dimensions of the particle are small compared with the wavelength of the incident radiation, this set can be represented as the oscillations of a single electric dipole. The forced oscillations of the scattering particle will take place with the same frequency (same wavelength) as the oscillations in the incident wave. If the primary wave is plane-polarized the secondary scattered wave will also be plane-polarized and its intensity will moreover be different in different directions. The scattering of solar rays in the atmosphere is the source of scattered radiation transmitted by the sky and at the same time the cause of many optical phenomena in the atmosphere.

A. Molecular scattering of light. Rayleigh's theory. The foundations of the theory of light scattering were laid by Lord Rayleigh in connection with the question of the blue color of the sky. Later this theory was considerably revised and developed for application to the conditions prevailing in the real atmosphere.* We will confine ourselves to stating the main conclusions which are relevant to our problems.

Rayleigh's theory of molecular scattering is based on the following fundamental assumptions:

1) the dimensions of the scattering particles are small compared to the wavelength of the incident light;

2) the scattering particles are spherically symmetric with respect to their optical properties (are spherical);

* A rigorous exposition of this question will be found in K.S. Shifrin's monograph "Rasseyanie sveta v mutnoi srede" (Scattering of Light in a Turbid Medium), GTTI, Moscow 1951, and also in M.V. Volkenshtein's "Molekulyarnaya optika" (Molecular Optics), Gostekhizdat, Moscow, 1951.

3) the scattering particles and medium are nonconducting and do not contain free electric charges;

4) the dielectric constant ε of the scattering particles differs only slightly from the dielectric constant of the medium;

5) the particles scatter light independently of each other. They are separated from each other by distances greater than the wavelength.

The well-known formula for the scattering coefficient obtained under these assumptions is

$$\alpha^{\varphi}_{\lambda,\,\mathrm{mol}} = \frac{\pi^2 (n^2-1)^2}{2r^2\lambda^4 N}(1+\cos^2\varphi). \tag{23}$$

The coefficient $\alpha^{\varphi}_{\lambda,\,\mathrm{mol}}$ shows that part of the radiation incident upon a unit volume will be scattered in the direction φ and will cross a unit surface normal to this direction at the distance r from the scattering center.

In the formula given above n is the index of refraction for dry air and N the number of molecules per cubic centimeter (Loschmidt number).

We can obtain the fraction of radiation scattered in all directions by a unit volume by integrating expression (23) over a spherical surface of radius r. Since the element of a spherical surface is $rd\varphi r\sin\varphi dA$ (A being the azimuth), we have

$$\alpha_{\lambda,\,\mathrm{mol}} = \frac{\pi^2 (n^2-1)^2}{2r^2\lambda^4 N}\int_0^{\pi}(1+\cos^2\varphi)r^2\sin\varphi\, d\varphi\int_0^{2\pi} dA = \frac{8\pi^3 (n^2-1)^2}{3\lambda^4 N} \approx \frac{32\pi^3 (n-1)^2}{3\lambda^4 N} \tag{24}$$

(under the condition that $n+1\approx 2$ and $n^2-1\approx 2(n-1)$).

The quantity $\alpha_{\lambda,\,\mathrm{mol}}$ is the volume monochromatic attenuation coefficient due to molecular scattering.

Earlier (§ 2) in calculating the attenuation of solar radiation in the atmosphere we referred the coefficient of attenuation to a unit mass of the atmosphere ($m=1$). Let us determine in this connection the magnitude of the scattering coefficient $\alpha_{\lambda,\,\mathrm{mol}}$ calculated for a vertical column of the atmosphere of unit cross section. Introducing, as before, the height of the homogeneous atmosphere H, we can write this coefficient in the form

$$a_{\lambda,\,\mathrm{mol}} = \alpha_{\lambda,\,\mathrm{mol}} H. \tag{25}$$

Numerically it turns out that $a_{\lambda,\mathrm{mol}} = 0.00868\,\lambda^{-4}$; when the wavelength dependence of the refractive index n is allowed for $a_{\lambda,\mathrm{mol}} = 0.00821\,\lambda^{-4.05}$.

By analogy with the transmission coefficient of the atmosphere p_{λ}, let us introduce the coefficient of transmission due only to scattering, denoting it by q_{λ},

$$q_{\lambda} = e^{-\alpha_{\lambda,\,\mathrm{mol}} H} = e^{-a_{\lambda,\,\mathrm{mol}}}. \tag{26}$$

Let us compare the calculated values with data obtained from observations under conditions approaching very close to the theoretical ones.

Table 22 lists the calculated values of q_{λ} and the values of p_{λ} obtained from observations at a height of 4420 m. From these data one can see that in the wavelength region where absorption is absent and under the conditions of a pure atmosphere (no large particles), the theoretically calculated values of the attenuation of direct solar radiation due to scattering only ($a_{\lambda,\mathrm{mol}}$) are in good agreement with the observational data.

If one knows the value of q_λ, the averaged scattering index a_m and the averaged transmission coefficient q_m for the entire solar spectrum for any atmospheric mass m can be found from the relation

$$S_m = S_0 e^{-a_m m} = S_0 q_m^m. \quad (27)$$

These quantities, like the transmission coefficient p, obviously depend on the atmospheric mass m traversed by the rays (Table 23).

TABLE 22

Values of q_λ (transmission coefficient for the Rayleigh atmosphere) and p_λ (from data of observations at a height of 4420 m)

$\lambda\mu$	0.35	0.39	0.45	0.50	0.60	0.80	1.00	1.60
q_λ	0.551	0.685	0.812	0.874	0.938	0.980	0.992	0.999
p_λ	0;552	0.673	0.789	0.861	0.895	0.957	0.955	0.935

TABLE 23

Values of the atmospheric transmission coefficient

Mass		$\frac{1}{2}$	1	2	3	4	6	8	10
Feussner and Dubois	a_m	0.102	0.097	0.089	0.082	0.077	0.067	0 059	0.055
	q_m	0.094	0.907	0.915	0.921	0.926	0.935	0.942	0.947
Kastrov	a_m	—	0.099	0.088	0.081	0.075	0.066	0.060	0.054
	q_m	—	0.906	0.916	0.922	0.928	0.936	0.942	0.947

The following empirical formula was obtained on the basis of the data cited in Table 23:

$$q_m = q_{m-1} m^{0.018}. \quad (28)$$

Once the values of the atmospheric transmission coefficient are known it is possible to calculate the distribution of energy in the solar spectrum at different depths in the Rayleigh atmosphere. The results of such calculations are presented in Figure 44, which shows that the magnitude of the flux

TABLE 24

Distribution of energy among individual spectral regions

Region of radiation spectrum	m							
	0	1	2	3	4	6	8	10
Ultraviolet (%)	6.7	4.2	2.7	1.8	1.1	0.5	0.2	0.1
Visible (%)	46.8	45.8	44.8	42.0	40.8	36.5	33.2	30.3
Infrared (%)	46.5	50.0	53.5	56.2	58.1	63.0	66.6	69.6
λ_{max} ($m\mu$)	475	500	525	585	615	625	645	665

of direct solar radiation decreases as the path travelled by the sun's rays in the atmosphere increases (increasing m); the wavelength with the maximum energy is shifted toward the longer waves; further, as m changes the distribution of energy among the individual parts of the spectrum,

ultraviolet ($\lambda < 0.40\mu$), visible and infrared ($\lambda > 0.74\mu$), also changes. The relevant data are given in Table 24.

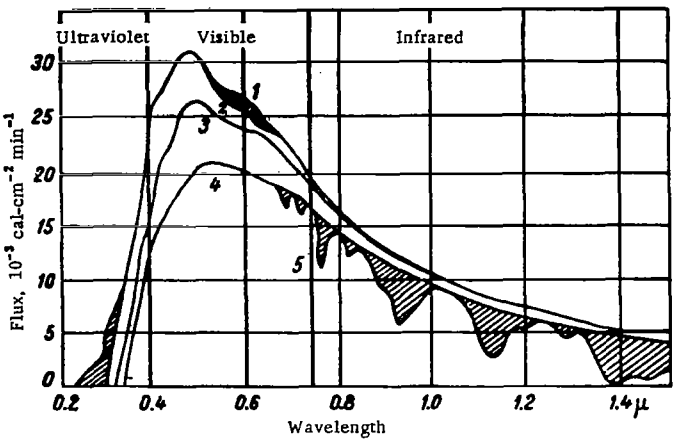


FIGURE 44. Change in the solar spectrum along the path through the atmosphere

1–spectrum outside atmosphere; 2–spectrum under ozone layer; 3–spectrum modified by scattering on air molecules; 4–spectrum due to influence of ozone and molecular scattering by aerosol particles; 5–absorption by water vapor (2 mm water layer).

It should be emphasized that all calculations from the above relations are applicable only to a certain idealized atmosphere, free from all large particles and nonabsorbing, which is generally called the ideal or Rayleigh atmosphere. Nonetheless they have proved very useful in explaining many phenomena and also in solving a number of practical problems.

Let us draw a few conclusions concerning this idealized atmosphere.

1. As one can see from formula (24), the monochromatic scattering coefficient $\alpha_{\lambda, \text{mol}}$ varies in inverse proportion to the fourth power of the wavelength. Correspondingly, in the first approximation (disregarding the dependence of n on λ) it decreases by a factor of 16 when the wavelength increases by a factor of two. For example, if one were to take $\alpha_{\lambda, \text{mol}} = 1$ for $\lambda = 0.7\mu$, for $\lambda = 0.30\mu$ $\alpha_{\lambda, \text{mol}}$ would be 16. Thus when the flux of direct solar radiation passes through the atmosphere rays with short wavelengths — violet and blue — are scattered much more strongly.

2. The scattering coefficient $\alpha_{\lambda, \text{mol}}$ depends on direction. This dependence — the so-called scattering indicatrix — is expressed by the factor $(1 + \cos^2\varphi)$. In Figure 45 the scattering indicatrix is shown by the solid outer curve. This curve is so constructed that the length of the radius vector drawn from the point representing the scattering center be proportional to the flux in the given direction. One can see that the scattering in the forward $(\varphi = 0)$ and in the backward direction $(\varphi = 180°)$ will be twice as large as the scattering in directions perpendicular to the incident light ($\varphi = 90$ and $180°$).

3. Scattering is directly proportional to the concentration of scattering particles, i.e., their number per unit volume. While this is not explicitly

shown by formula (24), in which N occurs in the denominator, it should be recalled that $(n^2 - 1)$ is proportional to N, whence this statement follows immediately.

4. The complete theory shows that the scattering of energy depends on the volume of the scattering particle v, increasing as the square of the volume or the sixth power of linear dimensions (radius). Thus when the radius of the particle increases by a factor of two the scattering coefficient increases by a factor of sixty-four.

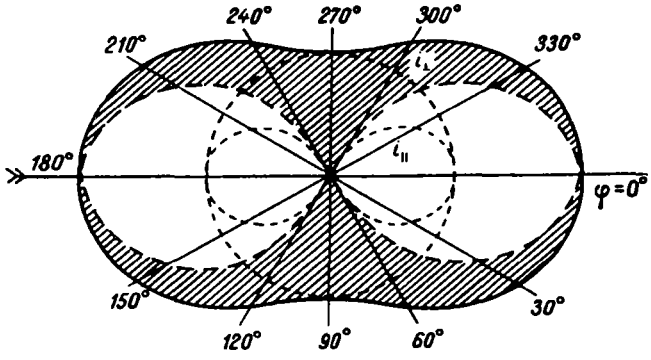


FIGURE 45. Molecular scattering indicatrix

5. The theory shows that polarization takes place in the scattering of natural light (for greater detail on this question see Chapter 24).

Finally, we note that the entire argument applies to single scattering only. In reality one finds multiple scattering, in which a ray scattered at some point experiences repeated scattering on other particles along its path. However, it is an extremely complicated matter to allow for this phenomenon, even for the restricted case of double scattering. Such allowance becomes necessary only when one is investigating the attenuation of light in very strongly scattering media.

B. Scattering of radiant energy by large particles. Mie's theory. Rayleigh's theory is applicable to particles with diameters not exceeding one-tenth of the wavelength. In reality the atmosphere always contains solid and liquid particles the dimensions of which are not only commensurate with the wavelength of incident light ($\lambda \approx 10^{-5}$ cm) but may even exceed it significantly. The regularities of light scattering on such aerosol particles differ essentially from those which follow from the theory of molecular scattering.

The general theoretical solution of the problem of scattering on aerosol particles the dimensions of which are commensurate with the wavelength was given by Mie. We will skip over the mathematical aspect of this theory in view of its complexity and confine ourselves to an exposition of its principal conclusions.

As Mie showed, the principal physical difference between scattering on small and large particles is the following. The field inside a particle can be regarded as uniform only when its dimensions are smaller than the wavelength of the monochromatic radiation incident upon it. The proper emission of the particle which arises in this process can then be regarded

as the emission by a dipole. However, if the particle dimensions are comparable with (or greater than) the wavelength of the incident light, the field around the particle can no longer be regarded as uniform and the spherical wave which it emits must be represented by a system of partial waves which become more complex as the particle dimensions increase.* Mathematically the intensity of the scattered light can then be treated as an infinite series the terms of which characterize the intensity of the indicated partial waves. These terms are expressed in spherical functions and contain coefficients having as arguments the quantities

$$\rho = \frac{2\pi r}{\lambda} \tag{29}$$

and

$$\beta = \frac{n}{n_0}\rho, \tag{30}$$

where r is the radius of the scattering particle and n and n_0 are respectively the refractive index of the particle and of air.

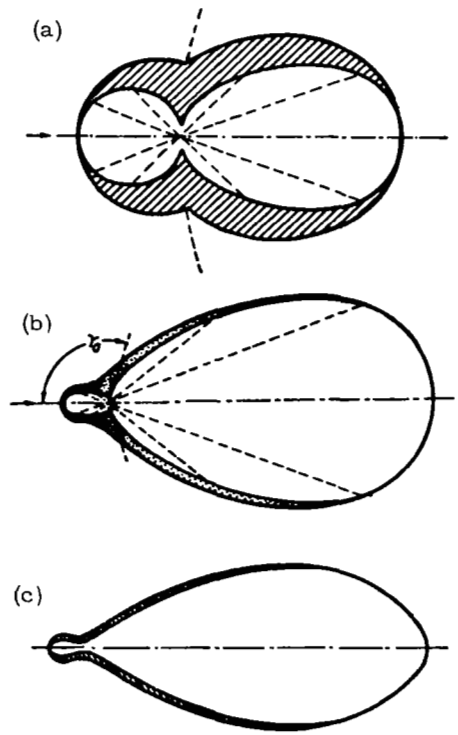


FIGURE 46

If natural light of unit intensity (wavelength λ) falls on a spherical particle of radius r, then the intensity of the light scattered in the direction

* On the dipole field are superposed fields of higher orders (quadrupole, octupole, etc.)

which makes an angle φ (scattering angle) with the direction of the primary ray at the distance d from the scatterer is given by $I=I_{\text{I}}+I_{\text{II}}$, where I_{I} and I_{II} are the intensities of the polarized components of the scattered light (I_{I} being the intensity of the component polarized in the plane of vision and I_{II} that of the component polarized in the perpendicular plane; (concerning the polarization of light see Chapter 24, § 5).

The calculations of greatest interest to us are those for a transparent sphere, which in relation to the atmosphere corresponds to the scattering of light on water droplets ($n=1.33$). Figure 46a shows the polar diagram of scattering for the case where $\rho=\frac{2\pi r}{\lambda}=1$, i.e., where $2r=0.32\ \lambda$ (after V. V. Shuleikin). As one can see, the scattering diagram for these particle dimensions is no longer symmetric like the Rayleigh indicatrix (Figure 45). In fact, it is elongated in the forward direction, so that the energy scattered forward exceeds the energy scattered backward by a factor of 2.37 and the energy scattered in directions perpendicular to the indicent ray by a factor of 2.85. The proportion of polarized light (cross-hatched part of graph) decreases simultaneously and in no direction is there complete polarization. As the particle dimensions increase these features become still more pronounced, as one can see from Figure 46b and c, which give the diagrams calculated by Shuleikin for the values $\rho=3(2r=0.96\lambda)$ and $\rho=9(2r=2.8\lambda)$.

This contrast in the intensities of scattered light in the forward and backward directions is called the "Mie effect".

The intensity of scattering increases very rapidly with particle size. Thus if the diameter of a particle is increased by a factor of ten, say from 50 to 500μ, then for $\varphi=180°$ the intensity of scattering will increase by about $5\cdot10^6$ times.

Since large particles are always present in the real atmosphere the observed curve for scattered light is the more complex elongate indicatrix and not the symmetric Rayleigh indicatrix $(1+\cos^2\varphi)$.

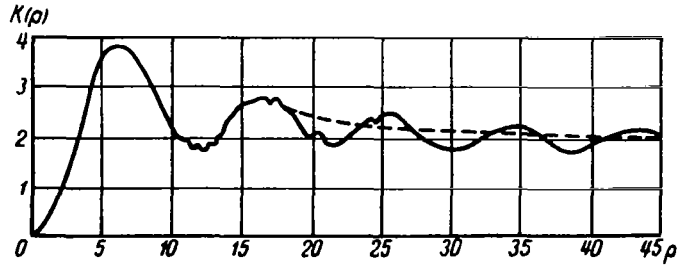


FIGURE 47. Scattering coefficient for water drops

The attenuation of the intensity of radiant energy due to scattering on large particles can be determined in the following way. If a spherical particle of radius r lies across the path of a parallel monochromatic beam the attenuation of the beam due to scattering on this particle will be proportional to its geometrical cross section πr^2, i.e., the scattering coefficient, which we will denote by α_{aer}, can be written as

$$\alpha_{\text{aer}}=\pi r^2 K(\rho). \tag{31}$$

The proportionality factor $K(\rho)$ [scattering area ratio] is a highly complex function of the parameter $\rho=\frac{2\pi r}{\lambda}$, i.e., it depends on the dimensions of the particle and wavelength of the incident light. The function $K(\rho)$ can be calculated from Mie's formulas and its form for a water droplet (refractive index $n = 1.33$) is shown in Figure 47. In this figure the values of $K(\rho)$ (ordinate axis) are plotted against the argument ρ (abscissa); also plotted on the axis of abscissas [sic] is a scale of particle radii for three wavelengths, making it possible to obtain the dependence of the attenuation coefficient α_{aer} on the particle dimension for a definite wavelength. The figure shows, first of all, that: 1) the $K(\rho)$ curve is of an oscillatory nature which is due to interference; 2) the curve has a series of maxima and minima; thus the first maximum is at $\rho = 6.1$, after which the curve drops to the first minimum at $\rho = 11.2$. This dip indicates that scattering in this section— as in other analogous sections— becomes stronger for the longer waves than for the short waves: 3) as ρ increases the oscillations weaken and $K(\rho)$ approaches the value $K(\rho) = 2$ for extremely large particles (for $r\to\infty$). Hence follows the seemingly paradoxal fact that the scattering coefficient of such an extremely large particle is $2\pi r^2$, i.e., such a particle scatters twice as much radiant energy as its transverse cross section (πr^2) receives from the incident flux of radiation. However, this can be explained by the fact that for such large particles an additional effect, that of diffraction, comes into play. Diffracted light propagated forward within a very small angle (of the order of $\frac{\lambda}{r}$) combines with the scattered light, and as a result the scattering coefficient is increased by a factor of two (to the value $2\pi r^2$).

All of the foregoing argument applies to scattering caused by a single particle. In the case of scattering by many particles of the same size (under the condition that individual particles scatter independently) the volume attenuation coefficient should be given as

$$\alpha_{aer} = \pi r^2 N K(\rho), \tag{32}$$

where N is the number of particles per unit volume.

If the particles are of different sizes then

$$\alpha_{aer} = \Sigma_r \pi r^2 N_r K_r(\rho). \tag{33}$$

All the data given apply to drops of water. However, the atmosphere also contains dust particles which in theoretical calculations can be identified with opaque ideally reflecting spheres, for which the refractive index is $n=\infty$. The character of the $K-\rho$ curve for spherical dust particles is the same as for water droplets but its maxima and minima are less sharp. Altogether the curve is strongly shifted toward the smaller values of the parameter ρ, i.e., toward the longer waves. The character of the scattering indicatrix for solid particles remains basically the same as for water droplets.

Since under actual conditions particles of strongly contrasting nature and sizes are always present in combination, the above relations are seldom applicable, and special methods based on both theoretical and experimental results must be used to calculate the attenuation of light in the atmosphere due to scattering.

It is natural to enquire whether one could represent the attenuation

coefficient α_{aer} by an expression analogous to that for the Rayleigh scattering coefficient:

$$\alpha_{\lambda,\ aer} = c\lambda^{-b}, \tag{34}$$

where c and b are certain empirically determined constants with different values for different particle dimensions and different conditions.

The values of the constants c and b have been calculated by a number of authors for different natural conditions. If for molecular scattering $\alpha_{\lambda,mol}=0.0082\ \lambda^{-4.05}$, then for high-mountain conditions, at low vapor content and λ lying in the interval from 0.36 to 1.00μ it seems that $a_\lambda = 0.012\ \lambda^{-3.68}$, while for an atmosphere containing water vapor in the amount $w = 1$ cm of condensed moisture $a_\lambda = 0.008\ \lambda^{-1.75}$.

Thus scattering on large particles very substantially alters the comparatively simple relations of molecular scattering. Basically the presence of large particles in the atmosphere displaces the position of the peak in the spectrum of scattered light toward the longer waves and at the same time increases its intensity. This effect increases with the size of the scattering particles. As a result when a great number of large particles are present in the atmosphere molecular scattering can often be regarded as insignificant.

§ 5. Atmospheric transmission and its characterization. The transmission coefficient. The turbidity factor

To determine the atmospheric transmission quantitatively at different points and instants it is necessary to have convenient and easily determined characterizations. For a long time the main characterization of atmospheric transmission was the averaged transmission coefficient considered earlier:

$$p_m = \sqrt[m]{\frac{S_m}{S_0}}. \tag{35}$$

The transmission coefficient is sometimes defined as the ratio of the magnitudes of the fluxes of solar radiation for two values of the atmospheric mass differing by unity, from one of the following relations

$$p_m' = \frac{S_m}{S_{m-1}} \tag{36}$$

or

$$p_m'' = \frac{S_{m+1}}{S_m}. \tag{36'}$$

However, the numerical values of the transmission coefficient as determined from formulas (35)–(36') turn out to be somewhat different and moreover, as we noted above, dependent on the atmospheric mass traversed by the rays. Therefore, in order to obtain data that can be compared with each other, one must take values of the transmission coefficient calculated not only in the same way but also for the same masses, or else reduce them to the unit mass, i.e., to the value p_1. Various methods have been suggested for this reduction, but none has produced reliable results.

The numerical values of p_m as obtained from observations do not permit

exact estimation of the actual transmission of the atmosphere but are merely approximate characterizations of its turbidity. Even for the same mass the observed values of p_m vary considerably in time owing to frequent and irregular fluctuations of turbidity; its mean annual values also fluctuate over a broad range. Thus at Pavlovsk for a mean value of 0.745 over a period of 31 years (1906–1936) the extreme mean annual values of p_1 were 0.570 in 1912 and 0.765 in 1926. The annual march of p_1 may also vary slightly from year to year. In most places minimum values of the transmission coefficient are observed in the summer months when the atmosphere contains a large amount of water vapor and convection is stronger, while maximum values are observed in winter. As one might expect, its values increase with elevation above sea level; at a height of about 3.5 km, for example, the values $p_1 = 0.862$ have been observed. We recall for comparison that for the "ideal" atmosphere (see Table 23) the transmission coefficient $q_1 = 0.907$.

The difficulties involved in using the coefficient p_m to characterize atmospheric turbidity led to the suggestion of other parameters. The aim was to obtain a quantity independent of the mass crossed by the rays and at the same time reflecting the influence of the chief factors responsible for fluctuations of transparency (primarily the content of water vapor and dust). The parameter which gained widest currency was suggested by Linke and is known as the turbidity factor T. Its meaning will become clear from the following. As we saw, attenuation by the atmosphere can be regarded as the result of: 1) molecular scattering, 2) scattering and absorption of radiation by water vapor, 3) scattering and absorption of radiation by dust. Accordingly, the optical thickness of the atmosphere in the vertical direction can be expressed as the sum

$$\theta_\lambda = \int_0^\infty k_{\lambda,\,\mathrm{mol}}\rho dh + \int_0^\infty k_{\lambda,\,w}\rho_w dh + \int_0^\infty k_{\lambda,\,\mathrm{aer}}\rho_{\mathrm{aer}} dh, \qquad (37)$$

where $k_{\lambda,\,\mathrm{mol}}$ is the mass coefficient of molecular scattering, ρ the density of air, $k_{\lambda,\,w}$ the mass coefficient of attenuation (absorption) of radiation by water vapor, ρ_w the density of water vapor, $k_{\lambda,\,\mathrm{aer}}$ the mass coefficient of attenuation of radiation by aerosol particles and ρ_{aer} the concentration of the latter.

The monochromatic turbidity factor is the name applied to the quantity

$$T_\lambda = \theta_\lambda \left[\int_0^\infty k_{\lambda,\,\mathrm{mol}}\rho dh \right]^{-1}. \qquad (38)$$

Further, since θ_λ and $\int_0^\infty k_{\lambda,\,\mathrm{mol}}\rho dh$ can be regarded as coefficients of attenuation of radiation calculated for the entire vertical column of air, one can interpret T_λ as the number of dry and pure atmospheres for which the attenuation of solar radiation would be equal to the attenuation in the real (moist and dusty) atmosphere. Obviously, the values of T_λ are always greater than unity.

Introducing the turbidity factor T_λ for the attenuation of the monochromatic flux of direct solar radiation for a certain atmospheric mass m, we can write

$$S_{m,\,\lambda} = S_{0,\,\lambda} e^{-\theta_\lambda m} = S_{0,\,\lambda} e^{-T_\lambda \int_0^\infty k_{\lambda,\,\mathrm{mol}}\rho dhm}, \qquad (39)$$

and since $e^{\int_0^\infty k_{\lambda,\,\mathrm{mol}}\rho dh} = q_\lambda$ is the monochromatic transmission coefficient of the ideal atmosphere, we also have

$$S_{m,\lambda} = S_{0,\lambda} q_\lambda^{T_\lambda m}. \tag{40}$$

Integrating over all wavelengths, we obtain the following expression for the total flux of direct solar radiation

$$S_m = \int_0^\infty S_{0,\lambda} q^{T_\lambda m} d\lambda. \tag{41}$$

Introducing the values of q_m and T averaged over the entire spectrum, where T is the integral turbidity factor, we obtain

$$S_m = S_0 q_m^{mT}, \tag{42}$$

whence

$$T = \frac{1}{m \ln q_m} \ln \frac{S_m}{S_0}. \tag{43}$$

Since $S_m = S_0 p_m^m$, to determine T one can also make use of the formula

$$T = \frac{\ln p_m}{\ln q_m} \left(\text{or} \quad \frac{\lg p_m}{\lg q_m} \right), \tag{44}$$

where p_m is the integral transmission coefficient for the mass m.

At times one distinguishes between the turbidity factors for the long-wave ($\lambda > 0.625\mu$) and short-wave ($\lambda < 0.625\mu$) regions of the solar spectrum.

To calculate T from the above formulas one must employ the value of the solar constant S_0 and theoretical values of q_m given in Table 22.

The turbidity factor, like the transmission coefficient, depends to some extent on the traversed mass m; however, this dependence is distinctly weaker and the turbidity factor is a more explicit characterization of atmospheric transmission.

The values of the turbidity factor T vary over a broader range than the values of p. In addition to irregular fluctuations, a definite daily and annual march can be detected. Usually the largest values of T are observed in the early hours of the afternoon and, in the annual march, in the summer months (June–July). For example, according to data of observations at Pavlovsk, $T = 3.15$ in July and 1.81 in December for an annual mean of 2.52.

It is possible to detect a relation between T and latitude: T decreases with increasing latitude. Also apparent is a dependence on altitude; T decreases with increasing elevation above sea level.

In an attempt to circumvent the mass dependence of T Linke proposed replacing it by a new turbidity factor, which we will designate by T_L. Unlike T, the factor T_L refers not to the pure and dry atmosphere but to a pure and moist one containing one gram of water vapor in a column of unit cross-section (1 cm^2). T_L can be calculated from the same formulas as T but all theoretical values for the ideal dry atmosphere appearing in these formulas must be replaced by the corresponding values for the ideal but moist atmosphere with vapor content $w = 1\ g/cm^2$.

The turbidity factor T_L no longer displays any significant dependence on atmospheric mass and its variations reflect actual changes in atmospheric

transmission more closely than the variations of T. Calculations show that the integral turbidity factor T_L is roughly twice smaller than T, while in the short-wave emission region ($\lambda < 0.625\mu$) it is slightly greater than T $\left(\frac{T_L}{T} \approx 1\right)$.

It is very convenient to characterize the respective influences of water vapor and dust on the attenuation of radiation by means of the turbidity factor T. Indeed, from the fact that for a pure and dry atmosphere $T = 1$, the turbidity factor can be expressed as follows:

$$T = 1 + W + R,$$

where W is the so-called moist turbidity factor and R the residual turbidity factor.

The quantity W characterizes the influence of absorption and scattering by water vapor on the attenuation of solar radiation; the quantity R gives the effect of attenuation due to absorption and scattering by aerosol particles, i.e., dust and products of vapor condensation (water droplets, ice crystals).

§ 6. Reflection of solar radiation. The albedo

The solar radiation which penetrates to the earth's surface is partially reflected by it. Reflection also takes place in the atmosphere, chiefly from the surface of clouds. The amount of energy reflected by a certain surface depends on the properties and state of this surface. The reflecting power of any surface can be characterized by its albedo, which is the ratio of the flux reflected by the given surface in all directions to the flux of radiant energy incident on this surface. The albedo is usually expressed in percents.

The albedo of natural surfaces varies considerably with the wavelength of incident radiation. When studying the albedo in detail one considers not only the integral (total) albedo A for the entire spectrum of incident radiation but also the spectral albedo A_λ or albedo in individual, more or less broad regions of the spectrum (e.g., visible region). One can also consider individually the albedo for direct and scattered radiation.

TABLE 25

Albedos of various surfaces

Type of surface	Albedo (%)	Type of surface	Albedo (%)
Chernozem, dry	14	Fallow, dry	8–12
Chernozem, wet	8	Fallow, wet	5–7
Serozem, dry	25–30	Plowed field, wet	14–17
Serozem, wet	10–12	Rye and wheat in various	
Blue clay, dry	23	growth phases	10–25
Blue clay, wet	16	Fresh green grass	26
Yellow sand	35	Sun-dried grass	19
Gray sand	18–23	Forest vegetation (spruce,	
River sand	43	pine, oak)	10–18

We will confine ourselves for the most part to considering the integral albedo for total radiation. A few results of measurements by ground observation for very small sections of the underlying land surface are given in Table 25. From the table one can see that while the albedo of different surfaces varies within comparatively narrow limits (10–30%) it is strongly dependent on the nature of the cover and state of the surface. The wetness of the surface has an especially pronounced effect: the greater the humidity, the smaller the albedo. This is of great significance for the thermal regime of irrigated fields.

Measurements from aircraft or balloons make it possible to determine the value of the albedo for extensive areas. The following mean values of the albedo (%) were obtained by L. I. Zubenok in observations from heights of 50–400 m:

Coniferous forest	12±2
Deciduous forest	15±2
Meadow	19±2

The albedo of snow mantles is considerable and averages about 60%. Further, it varies strongly depending on the character of the snow surface. Thus for dry freshly fallen snow the albedo approaches 90% and can even be greater, while for dirty wet snow it drops to 20–30%. The albedo of ice covers has yet to be investigated in detail but is also highly variable, amounting, for instance, to 30–40% for sea-ice surfaces.

Observations have established the existence of a daily march of the albedo, characterized in most cases by an increase in the albedo with decreasing solar elevation.

Despite its importance the question of the albedo of water surfaces such as seas, lakes and other bodies of water, has not been sufficiently studied by observation. The available data indicate low values and an extreme variability of the albedo, especially as a function of the angle of incidence of the sun's rays and state of the water surface (absence or presence of waves). The same conclusion follows from theoretical calculations. For the direct solar radiation calculations give a value of 2% for large solar elevations (50–90°) on a smooth sea surface; but this value increases rapidly as the sun's height decreases, and thus for $h_{\odot} = 10°$ the sea's albedo is 35%, and for $h_{\odot} = 4°$ it is already 65%.

The values of the sea-surface albedo for scattered radiation are highly variable and depend on the distribution of the clouds over the sky and on their character. According to data of observations, the albedo of sea surfaces varies between 4 and 18%, 10% being the mean frequently assumed in calculations.

The dependence of the albedo of all surfaces on the height of the sun is responsible for the daily march, in which the minima are observed around midday. At the same time, the fact that the character of the underlying surface changes seasonally explains the significant annual march of the albedo. In middle and northern latitudes an increase in the albedo is usually recorded from the warm half of the year to the cold one. An idea of the variability of the albedo is provided by the following data: according to measurements by N. N. Kalitin at Pavlovsk, the ten-day mean varies between 20 and 41% through the year; at Tashkent, according to I. N. Yaroslavtsev, the annual mean varies between 19 and 29%.

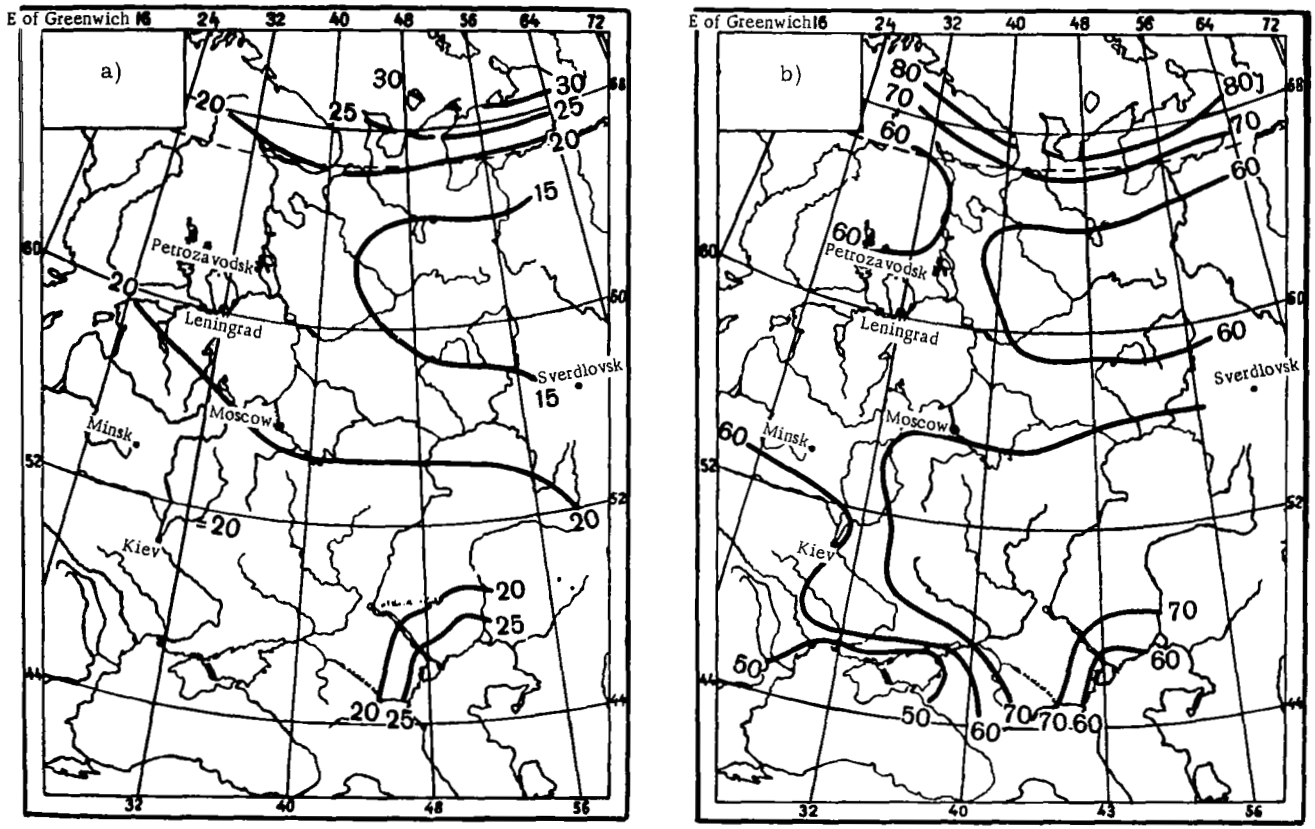


FIGURE 48. Albedo in the European USSR, summer (a) and winter (b) periods, after T.G. Berlyand

However, for many calculations it is important to know the mean values of the albedo, and that too for fairly large areas of the earth's surface and over more or less prolonged intervals in time (year, season, month, etc.).

Data of this kind make it possible to evaluate the albedo of various geographic zones and construct climatic charts. T. G. Berlyand has compiled seasonal and annual albedo charts for the European USSR based on her calculations. Two of these are given in Figure 48a (summer) and 48b (winter). As one can see winter values of the albedo are large throughout the territory in question and vary between 50% (in the south) to 80% (north). Summer values are considerably lower (20–30%), and the minimum values (about 15%) obtained in the northern districts of the European USSR and Urals, where large tracts are covered by coniferous forests.

Let us now consider the question of the reflection of solar radiation by clouds.

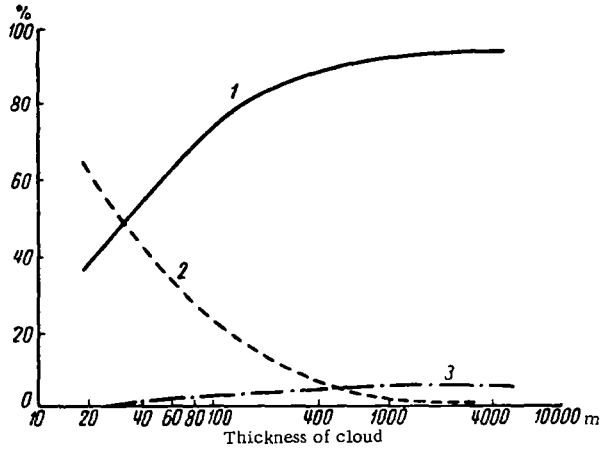


FIGURE 49. Dependence of reflection (1), transmission (2) and absorption (3) of direct solar radiation on cloud thickness

Measurements of the albedo of clouds are possible only with aircraft and balloons. They show this to be dependent on the vertical thickness of the clouds, increasing as the latter increases. It is also heavily dependent on the cloud form (type), the largest values being usually observed for altocumulus and stratocumulus clouds. The numerical values of the albedo are fairly large and, for a cloud thickness of, say, 300 m, can vary within 46–73% for different cloud forms (according to data of N. I. Chel'tsov). Owing to the great variability of the cloud albedo it is difficult to indicate a mean value; usually this is taken to be 50–55%.

Attempts have also been made to solve this problem theoretically. Although the calculations in question are admittedly tentative, they provide a qualitatively correct picture of the observed relations. Figure 49 illustrates the obtained dependence of the reflection, transmission and absorption of solar radiation on cloud thickness. One can see from the figure that the amount of radiation absorbed by clouds is limited and varies only

slightly with cloud thickness, whereas the albedo is large and increases rapidly with increasing thickness; the transmission decreases rapidly with thickness and is practically zero for a cloud thickness of the order of 1000 m.

If one knows the albedo of clouds and of various kinds of underlying land and sea surfaces one can determine the albedo of the earth as a whole, a datum which is of considerable interest for climatological and astrophysical investigations. It is obvious that the albedo of the earth as a planet can be expressed as the sum $A = A_s + A_a + A_c$, where A_s, A_a, A_c are respectively the albedo of the earth's surface, atmosphere (characterizing the back-scattering of the sun's rays by the atmosphere) and clouds.

The accepted value was previously $A = 42\%$ but revised calculations show this to be slightly too high. The value $A = 40\%$ is considered more accurate; of this the greatest part is due to the cloud albedo A_c (roughly 75 %) and the albedo of the earth's surface A_s is the least significant (about 7 %).

Chapter 9

SHORT-WAVE RADIATION IN THE ATMOSPHERE

Energy is delivered to the earth's surface at every instant by fluxes of solar radiation. If the magnitudes of these fluxes were known one could compute the total amount of energy received from each individual flux and from all fluxes together at any place in any time interval (hour, month, day, year). The theoretical treatment of this and many other actinometric questions presents very grave difficulties and is not always possible. As a result many of these questions can be answered only by direct measurements, to which a good deal of attention is devoted in actinometry. However, as this can be done only for a few isolated localities, indirect methods must be devised to calculate the energy received in the form of direct and diffuse radiation. In the present chapter we will consider the principal regularities established for direct, diffuse and total radiation by observation and by computations of various kinds.

§ 1. Direct solar radiation at the earth's surface

Let us first consider the distribution of direct solar radiation over the earth's surface and its variation in time. This problem is very simple to solve if one neglects the influence of the atmosphere. Indeed, in this case the amount of energy incident per square centimeter of horizontal surface at any point per unit time is

$$\frac{dQ}{dt} = \frac{S_0}{\rho^2} \sin h_\odot, \tag{1}$$

where S_0 is the solar constant, ρ the earth's distance from the sun, expressed in fractions of the mean distance, and $h_\odot$ the sun's height above the horizon at the given instant.

The value of $h_\odot$ at any point depends on the geographic latitude φ, time of year and time of day. From astronomy it is known that

$$\sin h_\odot = \sin \varphi \sin \delta + \cos \varphi \cos \delta \cos \tau, \tag{2}$$

where δ is the sun's declination, which changes over the year within $\pm 23.5°$ τ the hour angle, T the period of rotation of the earth about its axis (day) and t the time reckoned from noon.

Consequently, instead of (1) one can write

$$\frac{dQ}{dt} = \frac{S_0}{\rho^2} (\sin \varphi \sin \delta + \cos \varphi \cos \delta \cos \tau). \tag{3}$$

The above expression makes it possible to determine the amount of heat received from the sun at different geographic points in specified time intervals. Let us consider the heat influx from the sun per day at a certain point on individual days of the year and its distribution over the earth's surface on these days.

Obviously, the influx will be greatest at points for which $\tau = 1$ and simultaneously

$$\sin\varphi \cdot \sin\delta + \cos\varphi\cos\delta = \cos(\varphi - \delta) = 1. \qquad (4)$$

These are the points at which the sun is at the zenith $(\sin h_{\odot} = 1)$.

The smallest values (zero) are obtained for $h_{\odot} = 0$ when the sun is on the horizon.

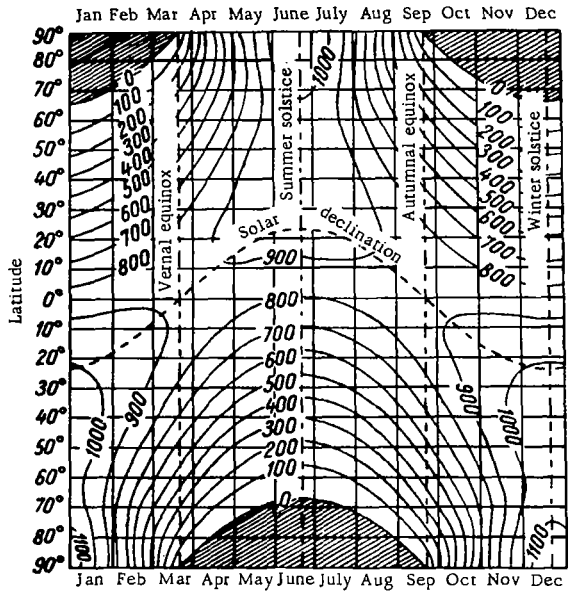

Jan Feb Mar Apr May June July Aug Sep Oct Nov Dec
Latitude
Vernal equinox
Summer solstice
Autumnal equinox
Winter solstice
Solar declination
Jan Feb Mar Apr May June July Aug Sep Oct Nov Dec

FIGURE 50. Latitude distribution of the solar radiation (incident in one day on one square centimeter of horizontal surface) in the absence of the atmosphere

Since at any definite point of the earth's surface ($\varphi = \text{const}$) the quantity δ changes very slowly, one can treat it as constant during one day and read its mean value from astronomical almanacs. One can then easily obtain the total amount of heat Q_d arriving at any point per square centimeter of horizontal surface at the upper boundary of the atmosphere (or on one square centimeter of the earth's surface in the absence of the atmosphere). To do this it is necessary to integrate expression (3) from sunrise to sunset (between the hour angles $\pm\tau$), treating δ and ρ as constant. We then obtain

$$Q_d = \frac{T}{\pi}\frac{S_0}{\rho^2}(\tau\sin\varphi\sin\delta + \cos\varphi\cos\delta\sin\tau). \qquad (5)$$

The results of calculations with formula (5) for an entire year are represented graphically in Figure 50; Figure 51 shows the annual march of insolation at different latitudes.

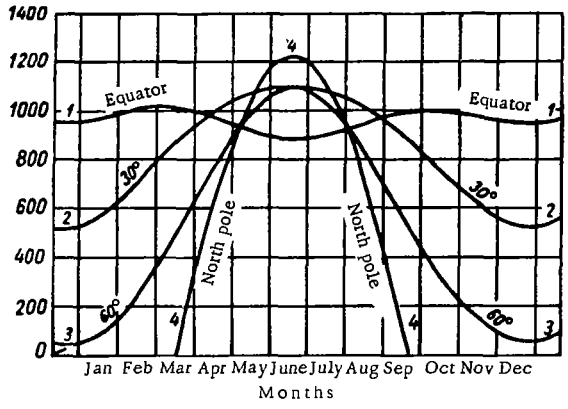


FIGURE 51. Annual march of insolation at different latitudes

From these figures one can see that near the equator (to latitude ± 15°) the daily heat influx displays two maxima, while farther poleward the march is simple with maximum at the solstice. Beyond the arctic circle the influx of radiation is zero for a certain period. It is interesting to note that in summer (10 May to 3 August) the top of the atmosphere receives more energy per day in the northern latitudes than at the equator. Thus at the summer solstice the pole receives 1.365 times more heat than the equator.

By continuing these calculations one can find the total amount of energy received by any horizontal sector of the earth's surface over different intervals in time. The distribution of such theoretical sums of heat over the earth's surface characterizes the so-called solar climate. The most complete characteristics of the solar climate are found in Milankovich's researches. However, the theoretical picture is very far from the actual one as it fails to account in any way for the influence of the atmosphere. To allow for this influence we introduce the transmission coefficient p_m into the initial expression (1) as a factor. Then for the energy flux we can write, instead of (3),

$$\frac{dQ}{dt} = \frac{S_0}{\rho^2} p_m^m (\sin\varphi \sin\delta + \cos\varphi \cos\delta \cos\tau). \qquad (6)$$

When calculating the daily sums Q_d from this formula one must take into account both the variation of the atmospheric mass m crossed by the rays and the variability of the transmission coefficient p_m. As a first approximation one can take a certain mean value p_m. Such calculations have been carried out for various values of the transmission coefficient p_m (0.6, 0.7, 0.8, 0.9, 1.0), and the results are presented in Figure 52. From the figure one can see how strongly the varying degrees of atmospheric transmission affect the heat influx. Also indicated are the semi-annual sums of heat (winter and summer); the magnitude of the contrast in heat influx between different latitudes during these periods is evident.

The quantities considered above give the so-called possible sums of heat. More accurate data for individual points can be obtained by taking different values of p_m — varying, also, over the year — on the basis

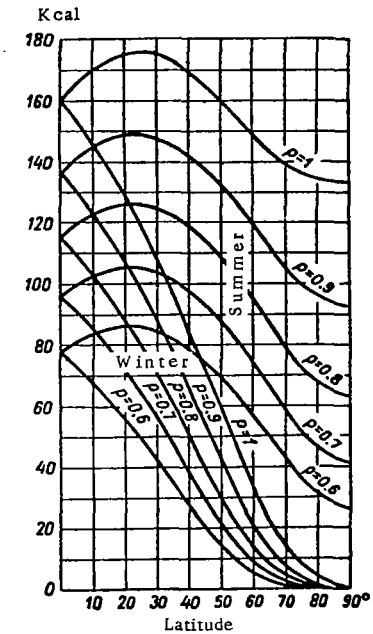


FIGURE 52

of observations or indirect considerations. But even so these possible sums will be remote from the actual ones, since the influence of clouds will not have been considered in their calculation. Their influence can be taken

TABLE 26

Annual sums of direct solar radiation (kcal) at various latitudes

	Latitude						
	60°N	50°N	40°N	30°N	20°N	0°N	20°S
Sum of radiation:							
theoretical	183	220	254	282	303	321	303
possible	105	128	148	163	172	187	172
actual	41	54	74	91	100	82	88
Ratio:							
actual to possible (%)	41	42	50	56	58	44	51
actual to theoretical (%)	22.5	25	30	32	33	25.5	29

into account only on the average and that by indirect methods. One will than have calculated the so-called actual sums of heat carried by the

direct solar radiation. More exact determination of the actual heat influx is possible only on the basis of observations.

To give an idea of the relations obtained when different methods are used to allow for the atmosphere's influence, we give the annual march of the daily sum of heat at Pavlovsk in Figure 53, as follows: theoretical sums in the absence of the atmosphere (curve A), theoretical sums for the ideal atmosphere (curve B), possible sums (curve C) and observed sums (curve D).

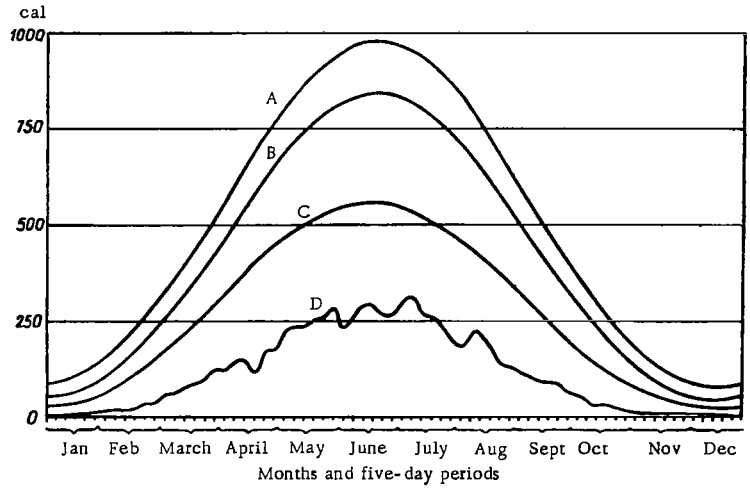


FIGURE 53. Annual march of daily sums of solar radiation on a horizontal surface, by five-day periods

By way of supplement to Figure 53 we present in Table 26 the annual sums of radiation in kilocalories according to latitude. These data show that under actual conditions low latitudes (near the equator) receive considerably more heat than high latitudes throughout the year; further, the greatest amount of heat is received not at the equator but at latitudes of about 20°. The table also shows how little of the theoretical and possible sums of energy (of the order of 25–30 and 50% respectively) actually penetrates to the surface. We note that even the energy which reaches the earth's surface is far from being fully utilized by it, as some of this energy is reflected and sent into interplanetary space.

§ 2. Direct solar radiation according to observations and calculations

The flux of direct solar radiation, measured at normal incidence, is conventionally denoted by S. The flux on a horizontal surface, denoted by S', is easily calculated from the relation

$$S' = S \sin h_{\odot}, \tag{7}$$

where $h_{\odot}$ is the sun's elevation above the horizon.

From the previous discussion we know that the magnitude of the direct flux at any point of the earth's surface at a certain instant, given by the relation $S=\frac{S_0}{\rho^2} p_m^m$, depends on the earth's distance from the sun ρ, on the atmospheric mass traversed by the rays m (i.e., on the solar elevation $h_\odot$) and on the state of the atmosphere characterized by the transmission coefficient p_m. The dependence of S' on solar elevation is even greater than that of S.

Obviously, within the earth's atmosphere the value of S is always smaller than the value of the solar constant S_0, and as solar elevation changes diurnally as well as annually the largest values are observed for the greatest solar elevations $h_\odot$.

Indeed, the maximum values of the direct flux S are usually observed around the true solar noon. These maximum values differ only slightly from point to point and amount to 1.5–1 6 cal/cm^2 · min; this is about 75–80% of the flux of solar radiation at the top of the atmosphere.

It is interesting to note that the observed maximum values are nearly the same at all latitudes. Thus in polar regions despite low solar elevations they attain the same values as in middle and even equatorial latitudes. This can be attributed to an increase toward the equator of the water vapor content, which largely determines the transmission of the atmosphere. The contrast in degree of atmospheric transmission compensates for the influence of differing solar elevations. The same explanation can be given for the fact that in many places the maximum values of S are observed not in summer, when solar elevation is highest, but in spring and even winter, when atmospheric transmission is maximal. For example, at Pavlovsk, Yakutsk and many other points in the USSR peak values of S are observed in March-April, and at Tashkent in January.

Thus, at best the earth's surface receives only about 80% of the flux of solar radiation incident on the upper boundary of the atmosphere.

It is also interesting to compare the observed maximum values of S with the amount of incoming radiation for an ideal (pure and dry) atmosphere. It appears that at certain points, e.g., Pavlovsk (Leningrad area), Yakutsk and Tashkent, the observed values are respectively 89.95 and 92% of the ideal values, i.e., in exceptional cases the actual atmosphere is fairly close to the ideal one with regard to transmission.

Obviously, as one moves upward over the earth's surface the flux of direct solar radiation increases. This vertical increase in S is somewhat slower in the upper layers than in the lower ones, where most of the solid and liquid impurities are concentrated. On the basis of a number of observations it can be assumed that the vertical gradient of S is of the order of $10^{-2}-10^{-3}$ cal/cm^2 · min per 100 m. In the lower layer (roughly to 2000 m) it amounts to about 0.02 cal/cm^2 · min per 100 m. Measurements with sounding balloons show that at a height of 22.6 km the flux of solar radiation increases to 1.78 cal/cm^2 · min.

The annual and daily march of the flux of solar radiation on cloudless days (or in the presence of clouds which do not obscure the solar disk) is determined mainly by variations in solar elevation, a certain influence being exercised also by variations in the transmission of the air. As a result the annual march displays roughly the same character everywhere, with maximum in the late spring or early summer months and minimum in

December. The amplitude of annual variation is considerably greater for the flux on a horizontal surface.

As to the daily march, maximum values occur at noon and minimum values (zero) in the night hours (from sunset to sunrise).

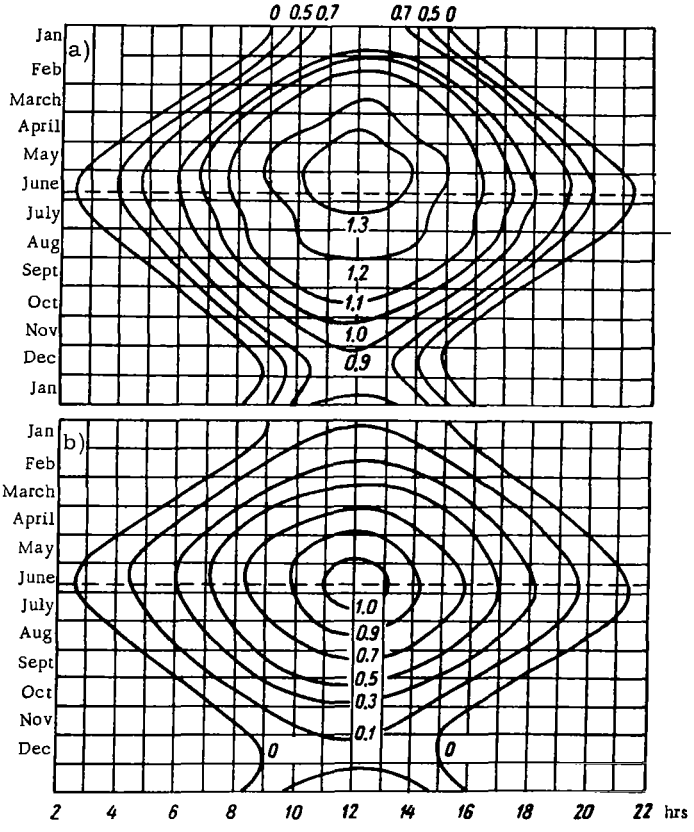


FIGURE 54. Isograms of direct solar radiation on normal (a) and horizontal (b) surface, Pavlovsk

Dashed line designates summer solstice.

A very clear representation of the annual and daily march is achieved by the construction of isograms, shown in Figure 54 for Pavlovsk (Leningrad area; mean for twenty years).

If data of continuous observations (or better yet recordings) are available for the flux of direct solar radiation, information can be obtained concerning the amount of energy (heat) furnished by this flux over a certain time interval (day, month, season, year). Data on the sums of heat of direct solar radiation are of great climatological and practical value. However, the number of points at which continuous registration of the flux of direct solar radiation by means of actinographs is being carried out is very

limited. This gave rise to the necessity of evolving indirect methods for obtaining the actual sums. In § 1 we showed how the possible sums of heat on a horizontal surface Q_p could be calculated. Since the actual sums of heat Q_a differ from Q_p due to the influence of clouds, in calculating the actual sums one must allow for clouds. Obviously, the general relation connecting Q_a and Q_p is

$$Q_a = Q_p f(a), \tag{8}$$

where $f(a)$ is a certain function of the parameter a which characterizes the influence of clouds; for a clear sky $f(a) = 1$.

Different methods of calculating the mean actual sum of solar radiation have been proposed by many authors (S. I. Savinov, V. N. Ukraintsev, S. I. Sivkov, et al.). We mention for illustration Savinov's formula for the mean monthly sum of solar radiation

$$Q_a = Q_p \frac{s_1' + 1 - \bar{n}}{2}, \tag{9}$$

where s_1' is the relative sunshine duration and $\bar{n}$ the mean amount of cloud, and also Ukraintsev's empirical formula

$$Q_a = a(s + b), \tag{10}$$

where a and b are constants and s is the monthly sum (in hours) of sunshine duration.

We note that the accuracy of results obtained by these indirect calculations depends on whether the choice of data on atmospheric transmission and clouds was an accurate and fortunate one. Comparison with data obtained by direct observation shows that mean annual sums calculated by these indirect methods are accurate to 5% on the average. The error exceeds 5% only in a few cases and is generally greater for winter than for summer.

TABLE 27

Annual sums of heat from direct solar radiation incident on horizontal surfaces $\Sigma S'$ kcal/cm² year

Observation point	Latitude (°)	$\Sigma S'$
Tikhaya Bay	80	14.4
Pavlovsk	60	39.8
Irkutsk	52	60.0
Vladivostok	43	60.0
Tbilisi	42	75.1
Tashkent	41	101.6

As to how much heat is delivered to the earth's surface by direct solar radiation, an answer is provided by Table 27, which lists data for a number of points in the USSR (cited in order of decreasing latitude).

From the table one can see that the sum of incoming energy increases on the whole with decreasing latitude; at the same time its values may be substantially different not only at different points but also at points situated at roughly the same latitude (cf data for Tbilisi and Tashkent).

This difference is also felt in the annual march. Figure 55 gives the annual march of possible sums (solid curves) for Pavlovsk (P), Tashkent (T) and Vladivostok (V), and that of actual sums (dashed curves) for the same points. Whereas the march of possible sums is nearly the same at all points, the march of actual sums is highly variable. This difference is due chiefly to different cloudiness conditions. At Vladivostok, for example, maximum cloudiness is observed in summer and therefore summer values of the actual sum are smaller than the spring and autumn ones.

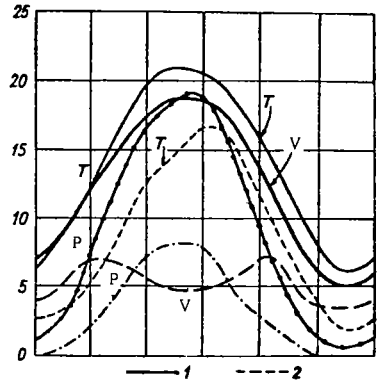


FIGURE 55. Annual march of possible (1) and actual (2) sums of heat at Pavlovsk (P), Tashkent (T) and Vladivostok (V)

It should be noted that owing to variations in atmospheric transmission and cloudiness from year to year the actual sums obtained in individual years — and all the more so in certain months of different years — may vary widely for the same point.

§ 3. Diffuse radiation

Diffuse radiation arrives at the point of observation from all points of the sky. One usually considers its flux on a horizontal surface, denoted by D.

The theoretical calculation of the intensity and flux of scattered radiation in general form is an extremely complicated problem. In the first approximation it can be solved for the case of the ideal atmosphere, where absorption can be disregarded.

Let us assume that the optical properties of the atmosphere are the same in all layers and that the latter differ only in the concentration of scattering particles. We will also disregard the effect of multiple scattering and the effect of reflection of radiation by the earth's surface. Furthermore, we will allow neither for the curvature of the earth's surface and atmosphere, nor for refraction.

Let us take (Figure 56) a volume element situated at the height h in this atmosphere, subtending a solid angle $d\omega$ at the point of observation O in the direction of the zenith angle z.

This volume, when irradiated by the sun's rays (zenith distance of the sun $z_\odot$), will scatter radiation in the direction r making the angle $(\pi - \varphi)$ with the incident rays. The intensity of this scattered radiation $dI'_{\lambda,\varphi}$ is, from Rayleigh's law, proportional to the volume dv and to the flux of solar radiation S_λ incident upon it; that is,

$$dI'_{\lambda,\varphi} = \alpha^{\varphi}_{\lambda,\,\mathrm{mol}} S_\lambda \, dv, \tag{11}$$

where $\alpha^{(\varphi)}_{\lambda\,\mathrm{mol}}$ is the volume molecular scattering coefficient in the direction φ, given by the relation

$$\alpha^{(\varphi)}_{\lambda,\,\mathrm{mol}} = \frac{\pi^2 (n^2-1)^2}{2\lambda^4 N}(1+\cos^2\varphi). \tag{12}$$

Bearing in mind that $dv = r^2 dr d\omega$, we rewrite (11) as

$$dI'_{\lambda,\varphi} = \frac{\pi^2 (n^2-1)^2}{2\lambda^4 N} (1+\cos^2\varphi) S_\lambda r^2 dr d\omega. \qquad (13)$$

We will take into account the attenuation of direct solar radiation on its way from the top of the atmosphere, where it is $S_{0,\lambda}$, to the volume in question, bearing in mind that in the absence of absorption it will be determined only by scattering.

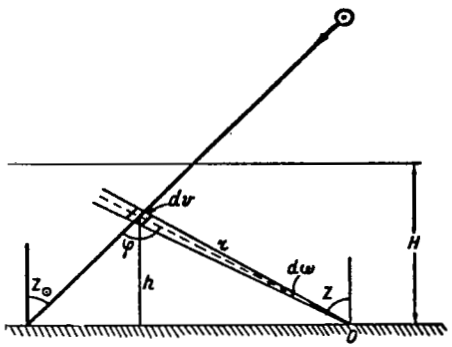


FIGURE 56

We denote by $k_{\lambda,\,\mathrm{mol}} = \frac{\alpha_{\lambda,\,\mathrm{mol}}}{\rho}$ the mass coefficient of molecular scattering. Here ρ is the density of air at the height h and $\alpha_{\lambda,\,\mathrm{mol}}$ the volume coefficient of molecular scattering in all directions. As we saw earlier (cf. formula (24) of Chapter 8),

$$\alpha_{\lambda,\,\mathrm{mol}} = \frac{8\pi^3 (n^2-1)^2}{3\lambda^4 N}$$

(henceforth we drop the subscript "mol").

In accordance with formula (27) of Chapter 8 we can write

$$S_\lambda = S_{0,\lambda} e^{-k_\lambda M m_\odot}, \qquad (14)$$

where M is the mass of air in a vertical column of unit cross section located at the level h and m the relative mass of the atmosphere $m_\odot = \sec z_\odot$ traversed by the sun's rays.

We will allow for the attenuation of scattered radiation along the path from the volume dv to the observation point in a similar manner, writing

$$dI_{\lambda,\varphi} = \frac{dI'_{\lambda,\varphi}}{r^2} e^{-k_\lambda M_h m}, \qquad (15)$$

where $dI_{\lambda,\varphi}$ is the intensity at the point O of radiation scattered from the volume in question, M_h the mass of air in a vertical column of air from the earth's surface to the level h and m the relative atmospheric mass traversed by the rays.

Thus, bearing in mind (13) and (14), we can write the following expression for the intensity $dI_{\lambda,\varphi}$ at the point O

$$dI_{\lambda,\varphi} = \frac{\pi^2 (n^2-1)^2}{2\lambda^4 N} S_{0,\lambda} (1+\cos^2\varphi) e^{-k_\lambda [M m_\odot + M_h m]} dr d\omega. \qquad (16)$$

Since $M=M_0-M_h$, where M_0 is the mass of the entire vertical column of air of unit cross section to the top of the atmosphere, and $dM=\rho dh=\rho\frac{dr}{m}$, whence $dr=\frac{mdM}{\rho}$, we finally obtain in place of (16)

$$dI_{\lambda,\varphi}=\frac{\pi^2(n^2-1)^2}{2\lambda^4N}S_{0,\lambda}(1+\cos^2\varphi)\,e^{-k_\lambda M_0 m}\,e^{-k_\lambda M(m_\odot-m)}\times\frac{mdM}{\rho}d\omega. \quad (17)$$

Integrating the above over M from O to M_0 and bearing in mind that $\frac{(n^2-1)^2}{N\rho}$ does not depend on M, we obtain an expression for the intensity of scattered radiation from the entire atmospheric cone within the solid angle $d\omega$ at the point O (setting $d\omega=1$)

$$\begin{aligned}I_{\lambda,\varphi}&=\frac{\pi^2(n^2-1)^2}{2\lambda^4N\rho k_\lambda}S_{0,\lambda}(1+\cos^2\varphi)\,e^{-k_\lambda M_0 m}\,m\int_0^{M_0}e^{-k_\lambda(m_\odot-m)}\times\\&\times dMd\omega=\frac{\pi^2(n^2-1)^2}{2\lambda^4N\rho k_\lambda}S_{0,\lambda}(1+\cos^2\varphi)\frac{m}{m_\odot-m}\times\\&\times\left[e^{-k_\lambda M_0 m}-e^{-k_\lambda M_0 m_\odot}\right],\end{aligned} \quad (18)$$

and since

$$\frac{\pi^2(n^2-1)^2}{2\lambda^4N}\frac{1}{k_\lambda\rho}=\frac{\pi^2(n^2-1)^2}{2\lambda^4N}\frac{3\lambda^4N}{8\pi^2(n^2-1)^2}=\frac{3}{16\pi},$$

we have

$$I_{\lambda,\varphi}=\frac{3S_{0,\lambda}}{16\pi}(1+\cos^2\varphi)\frac{m}{m_\odot-m}\left[e^{-k_\lambda M_0 m}-e^{-k_\lambda M_0 m_\odot}\right]. \quad (19)$$

Introducing the spectral transmission coefficient for the ideal atmosphere $q_\lambda=e-k_\lambda M_0$, one can write formula (19) as

$$I_{\lambda,\varphi}=\frac{3S_{0,\lambda}}{16\pi}(1+\cos^2\varphi)\frac{m}{m_\odot-m}\left[q_\lambda^m-q_\lambda^{m_\odot}\right], \quad (19')$$

which gives the intensity of monochromatic scattered radiation.

To obtain the total (integral) intensity of scattered radiation one must integrate expression (19') over all wavelengths from 0 to ∞, i. e., one must write

$$I'_\varphi=\frac{3}{16\pi}(1+\cos^2\varphi)\frac{m}{m_\odot-m}\int_0^\infty S_{0,\lambda}\times\left[e^{-k_\lambda M_0 m}-e^{-k_\lambda M_0 m_\odot}\right]d\lambda. \quad (20)$$

The same thing can be done for a definite region of the spectrum bounded by the wavelengths λ_1 and λ_2.

However, the integration can be carried out only numerically or graphically, due to the fact that $S_{0,\lambda}$ is a very complex function of the wavelength.

Recalling that the angle φ which appears above is related to z, $z_\odot$ and the azimuth ψ of the sighted point by $\cos\varphi=\cos z\cos z_\odot+\sin z\sin z_\odot\cos\psi$, one can find out how the intensity of scattered radiation will be distributed over the sky. The problem becomes considerably simpler if the sighted point lies in the sun vertical ($\psi=0°$) or in the antisolar direction ($\psi=180°$), where $\cos\varphi=\cos z\cos z_\odot+\sin z\sin z_\odot=\cos(z-z_\odot)$, or $\cos\varphi=\cos z\cos z_\odot-\sin z\sin z_\odot=\cos(z+z_\odot)$. Calculations of the integral

radiation in this case are represented in Figure 57 for different zenith distances of the sun (0, 30, 60, 75, 87°).

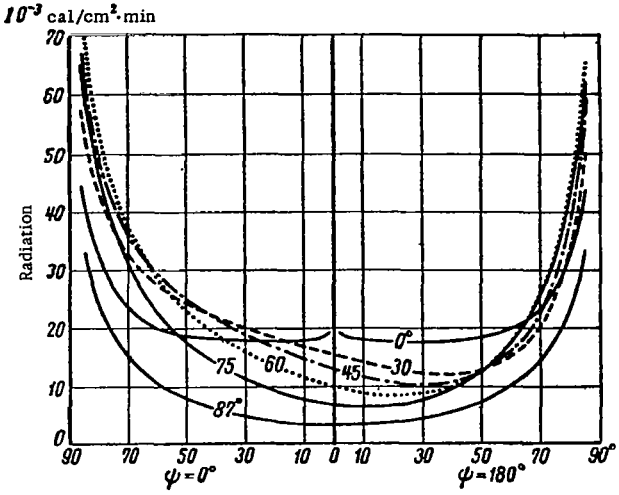


FIGURE 57. Distribution of integral intensity of scattered radiation in the sun vertical in the solar and antisolar directions

From the figure one sees that the integral intensity of scattered radiation in the sun vertical is comparatively small and changes only slightly near the zenith. It increases rapidly as one approaches the horizon and its peak intensity at the horizon occurs for $z_\odot = 60°$. A significant point is that the intensity of scattered radiation is somewhat greater in the sun vertical ($\psi = 0°$) than in the antisolar direction ($\psi = 180°$).

Making use of the relations (19) and (20) for the intensity of scattered radiation, one can obtain expressions giving the magnitude of the scattered radiation arriving from any section of the sky by integrating the expression for the intensity of scattered radiation over the solid angle subtended by the given section. It is simplest to do this for a circular zone of width dz located at the angular distance z from the zenith. Then, integrating over all solid angles $d\omega = \sin z\, dz\, d\psi$ located within the indicated circular zone, we obtain the so-called zonal radiation on a surface normal to the rays D'_{zon}. Multiplying D'_{zon} by $\cos z$, we obtain the zonal radiation on a horizontal surface $D_{zon,n}$

Figure 58 shows results of calculations of the zonal radiation on a horizontal surface for various values of the zenith distance of the sun ($z_\odot$ equal to 0, 30, 45, 60, 75 and 87°). As one can see, as the zenith distance increases the zonal radiation first rises (for any $z_\odot$) to a peak, then decreases. The maximum of zonal radiation is shifted toward the larger zenith distances with increasing zenith distance of the sun, while the height of the maximum decreases and the curve as a whole becomes flatter. The highest peak of zonal radiation from a zone 10° wide occurs at $z_\odot = 0°$ and $z_\odot = 47.5°$. By summing up the contributions of all circular zones of the sky one can calculate the flux of diffuse radiation D. This flux decreases

monotonically with increasing zenith distance of the sun, which accounts for the diurnal march with its maximum at true solar noon.

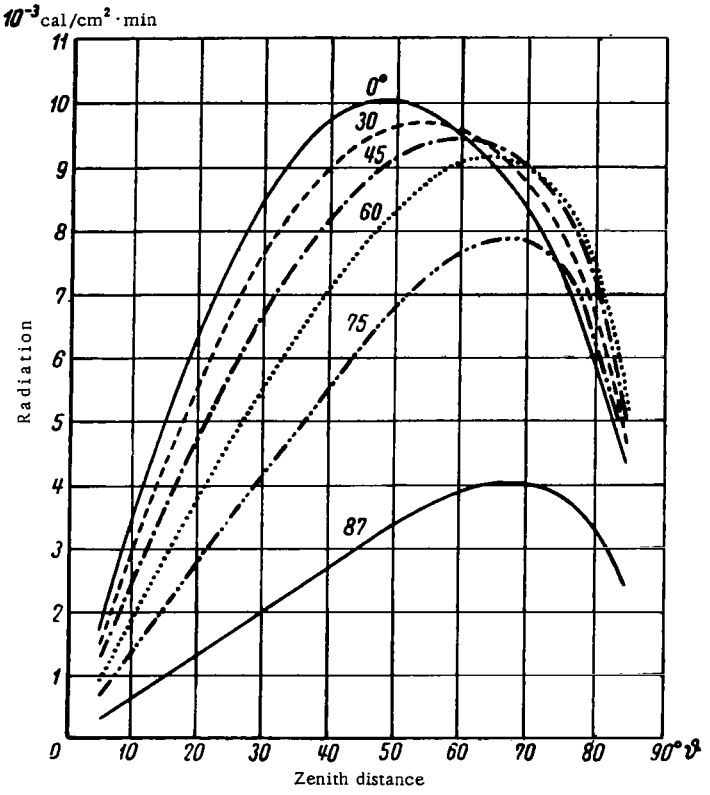


FIGURE 58. Zonal radiation incident on a horizontal surface

Let us compare the flux D with the flux of direct solar radiation under the same conditions (the ideal atmosphere). The ratio $\frac{D}{S}$ increases with increasing $z_{\odot}$, especially for large $z_{\odot}$. As for the total radiation $(D+S')$, in the ideal atmosphere it decreases monotonically with increasing zenith distance to the sun and therefore displays a simple daily march with peak at true noon.

The application of the obtained theoretical relations to the actual atmosphere is complicated by two factors; first, the presence inside it of an extremely wide variety of scattering particles, owing to which the scattering indicatrix becomes complicated and varies from point to point, and second, the necessity of accounting for absorption in addition to scattering. The problem of the scattering of radiation in the atmosphere is therefore highly complex. Important contributions to the theory of this problem have been made by Soviet scientists (V. A. Ambartsumyan, E. E. Kuznetsov, V. V. Sobolev, V. V. Shuleikin, V. G. Fesenkov et al.).

Another highly complex problem is that of the theoretical calculation of the scattering of radiation by clouds. We will mention the fundamental results without presenting the calculations. A result which agrees entirely

TABLE 28

Zonal radiation (%), from observations at Pavlovsk (Leningrad area)

$z_\odot$	Zones (°)								
	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90
75°	7,6	14,9	**17,0**	16,1	14,2	11,9	9,3	6,3	2,7
60	4,2	11,0	15,2	**17,9**	16,6	14,0	10,6	7,1	3,4
40	2,4	9,0	13,5	16,6	**18,4**	15,7	12,0	8,2	4,2

Remark: Boldface numbers indicate maximum values.

with the experimental data is that for not too thin stratiform clouds the scattered radiation emerging through the lower surface of the clouds is independent of the azimuth; on the other hand, for thin stratiform clouds the intensity of the scattered radiation is maximum in the zenith direction. The cloud albedo for scattered radiation increases with increasing cloud thickness while the transmission coefficient decreases, so that direct solar radiation does not pass through fairly thick clouds which, as has already been pointed out, transmit only scattered radiation through their lower surface.

TABLE 29

Flux of diffuse and total radiation

Atmospheric turbidity	Type of radiation	$h_\odot$					
		10	20	30	40	50	60
T = 1,0 (ideal)	D cal/cm²·min	0,028	0,053	0,074	0,091	0,102	0,108
	S' cal/cm²·min	0,229	0,524	0,812	1,083	1,316	1,507
	$\frac{D}{S'}$ %	12,2	10,1	9,1	8,4	7,8	7,2
T = 3,25 (turbid)	D cal/cm²·min	0,074	0,114	0,146	0,169	1,189	0,201
	S' cal/cm²·min	0,095	0,307	0,545	0,790	1,005	1,181
	$\frac{D}{S'}$ %	78	38	28	22	19	17
T = 5,50 (highly turbid)	D cal/cm²·min	0,113	0,167	0,208	0,239	0,260	0,273
	S' cal/cm²·min	0,040	0,180	0,366	0,561	0,784	0,925
	$\frac{D}{S'}$ %	283	92	57	42	34	30

Theoretical deductions are confirmed qualitatively by results of the observation of the flux of scattered radiation under a cloudless sky. Table 28 gives values of the flux of zonal radiation in percents of the radiation of the whole sky based on data of observations at Pavlovsk.

These data clearly show the indicated regularity of distribution of zonal radiation over the sky.

The integral flux of diffuse radiation from the whole sky D increases monotonically with increasing $h_\odot$ (decreasing $z_\odot$). Table 29 gives the

values of D, S' and the ratio $\frac{D}{S'}$ for the ideal atmosphere (first three rows), and the values of the same quantities as obtained on the basis of observations at Frankfurt-am-Main (Federal Republic of Germany) for a turbid ($T = 3.25$) and highly turbid ($T = 5.50$) atmosphere.

The data of this table show, in complete agreement with the theoretical results, that as $h_\odot$ increases the quantities D and S' increase, the increase in S' being considerably larger than the increase in D. As a result the relative contribution of D to the sum $D+S'$ is larger at low solar elevations and at $h_\odot = 0°$, i.e., sunrise and sunset, when the entire influx is due to scattered radiation. The influence of atmospheric turbidity is felt in the fact that with increasing turbidity the absolute value of D increases while S' decreases, i.e., for a general reduction of the sum $D+S'$ the contribution of diffuse radiation increases very strongly. From this one can see how effectively the influx of diffuse radiation compensates for the loss of direct solar radiation due to atmospheric attenuation.

The influence of clouds on the flux of diffuse radiation is particularly great. Clouds, which contain large scattering particles in great numbers, increase diffuse radiation considerably. This effect depends on the type of cloud, amount of sky covered by the cloud and solar elevation. Frequently diffuse radiation attains values comparable with the flux of direct solar radiation. At Pavlovsk, for example, the highest measured value of D over a twenty-year observation period was about 0.60 cal/cm^2 · min. High values of D are especially frequent in arctic regions where they often reach 1.0 cal/cm^2 · min.

TABLE 30

Flux of diffuse radiation (cal/cm^2 · min) depending on cloud form for different solar elevations $h_\odot$

Cloud form	$h_\odot$ (°)			
	5	10	20	40
Clear	0.03	0.05	0.08	0.11
Ci	0.05	0.09	0.16	0.26
Cs	0.06	0.11	0.19	0.34
Ac	0.06	0.11	0.22	0.39
As	0.05	0.09	0.17	0.30
Sc	0.01	0.04	0.13	0.27
St	0.02	0.04	0.08	0.16

The dependence of D on cloud type is apparent from Table 30, which cites mean values of D (cal/cm^2 · min) obtained for Pavlovsk. One can see that the strongest effect is due to middle clouds (Ac and As); high (Ci) and low clouds have less effect on the variation of D. As the sun's elevation increases the influence of clouds becomes stronger.

The flux of diffuse radiation is very strongly affected by the albedo of the earth's surface: the flux D increases perceptibly as the albedo rises. This is to be attributed to the multiple reflections experienced by the rays between the earth's surface and the lower edge of the clouds.

Under clear skies the annual and daily fluctuations of the flux D is fairly regular, as one can see from the graph in Figure 59; under conditions

of variable cloud the daily march is highly complex and irregular. For a continuous cloud cover the daily march is also comparatively simple; its maximum is recorded around midday.

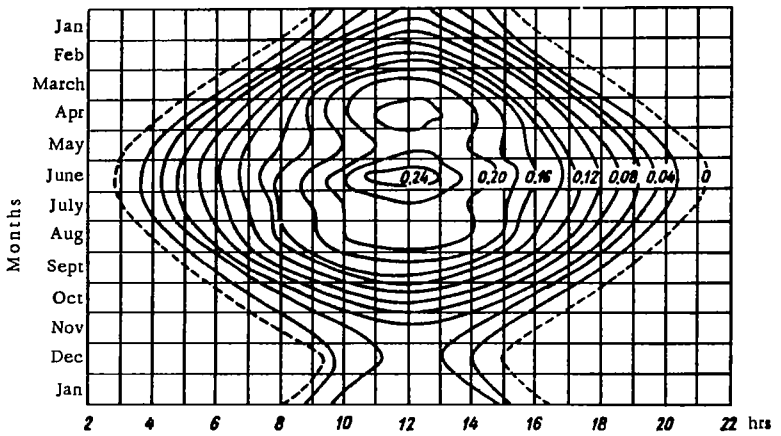


FIGURE 59. Isograms of diffuse radiation, Pavlovsk

As one ascends above sea level in a clear sky the flux of diffuse radiation falls off, which is understandable in view of the decreasing thickness of the scattering layers of the atmosphere. In the presence of clouds, however, this flux increases with height in the layers beneath the clouds.

The sums of heat delivered by diffuse radiation over a certain period can be found in the same way as for the direct solar radiation. To do this one can employ both results of continuous observations and empirical formulas.

TABLE 31

Mean daily sums of diffuse radiation (cal/cm^2 · day)

Observation point	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Year
Pavlovsk ($\varphi = 60°$)	24	63	116	149	179	195	188	164	113	55	22	13	107
Paris ($\varphi = 49°$)	40	67	109	168	215	244	234	191	135	85	48	35	131
Takubaya ($\varphi = 19°$, $H = 2300$ m)	88	82	186	252	268	201	135	202	156	117	135	141	164

Table 31 gives the average daily sum of diffuse radiation $q = \Sigma D$ (cal/cm^2) for each month of the year at several points. These data illustrate a definite annual march of diffuse radiation (maximum values in summer) and give a good idea of how much diffuse radiation contributes to the total influx of heat from the sun.

The sums of heat delivered per year to the earth's surface by diffuse radiation amount everywhere to several tens of percents of the sum of direct solar radiation. They are particularly large in high latitudes. In the Arctic the influx of heat due to diffuse radiation usually even exceeds (sometimes very considerably) that due to direct solar radiation. The role

of diffuse as against direct radiation is particularly great in winter. At Pavlovsk, for example, in December diffuse radiation delivers 3.5 times more heat to the earth's surface than direct solar radiation.

§ 4. Total radiation

Total radiation is the sum of the fluxes of direct and diffuse solar radiation incident upon a horizontal surface. By summing these fluxes over any period (day, month, year) one obtains the total sum of heat incident in this time on a unit surface.

The following symbols are in general use:

$\Sigma S' = Q$ — sum of direct radiation,

$\Sigma D = q$ — sum of diffuse radiation,

$(Q+q)$ — sum of total radiation.

The number of points at which direct and diffuse radiation are individually recorded is limited. It is much smaller than the number of points at which the total radiation is observed. For this reason the latter has been studied much more exhaustively, the more so as data on the total sum of energy delivered by solar radiation are more important in practical work. Correspondingly many more indirect methods have been proposed for calculating the latter.

The magnitude of the flux of total radiation and the regularities governing its variations are determined by the values of S' and D and their variations, which are controlled in turn by solar elevation, atmospheric transmission, cloudiness and other factors considered above. According to observational data, for clear skies the dependence of the total radiation on the sun's height is almost strictly linear and can be expressed by $Q+q=ch_{\odot}$, where, however, the empirical proportionality factor c is not rigorously constant; even for the same point it depends on atmospheric transmission. In the presence of clouds, of course, this relationship becomes far more complex.

TABLE 32

Annual march of mean daily sum of total radiation (cal/cm^2·day)

Observation point	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Year
Pavlovsk	25	70	194	269	341	398	445	296	176	69	19	11	193
Tashkent	65	119	186	293	332	541	522	511	387	225	124	58	280
Washington	173	243	337	418	484	513	491	438	357	292	205	155	344

In general, the amount, character and distribution of cloud over the sky are the most important factors controlling the flux of total radiation. Data cited earlier show that though the flux of diffuse radiation compensates partially for the attenuation of the flux of direct solar radiation S' in the atmosphere, this compensation is not complete. Therefore under a partly cloudy sky, provided the sun itself is not obscured by clouds, the flux of total radiation will be greater than under clear skies.

Turning to the sum of total radiation, we present in Table 32 the annual march of the mean daily sum at several points for purposes of illustration

At most points of the earth's surface the peak of the daily sum $(Q+q)$ occurs in the summer months, though at certain points the annual march may be complicated by the annual march of cloudiness. Indeed, the daily sum of total radiation is usually 2–3 times smaller under a continuous cloud cover than for clear skies. Thus summer values of the influx of total radiation may be exceeded by the spring and autumn values at points where the peak cloudiness is observed in the summer months (e. g., Vladivostok).

TABLE 33

Annual sum of solar radiation (kcal/cm^2·annum; direct Q, scattered q and total $Q+q$) for several points in USSR

Point	Q	q	$Q+q$	$\frac{q}{Q+q}$ %	$\frac{q}{Q}$ %
Tikhaya Bay ($\varphi = 80°$)	16	39	55	71	244
Pavlovsk ($\varphi = 60°$)	41	30	71	42	73
Karadag ($\varphi = 45°$)	80	37	117	32	46
Kislovodsk ($\varphi = 44°$)	66	53	119	44,5	80
Tashkent ($\varphi = 41°$)	100	37	137	23	37

Naturally, the annual and monthly influx of total radiation varies widely from point to point depending on latitude, cloudiness and atmospheric

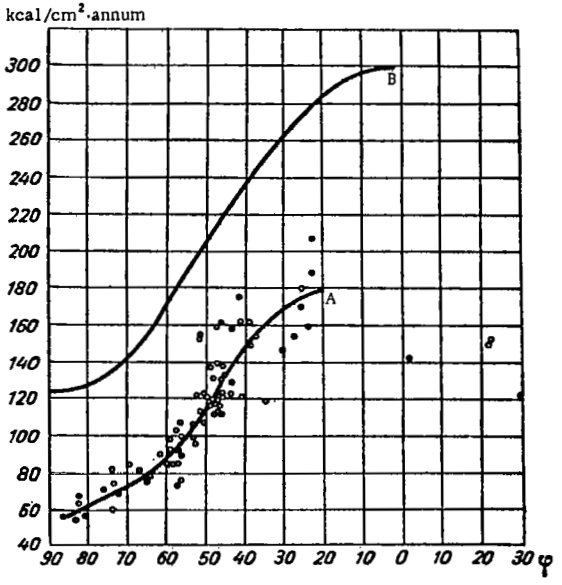


FIGURE 60. Variation of mean annual total radiation (A) and radiation at top of atmosphere (B) with latitude

transmission. The roles of the individual terms of this sum — the direct and the diffuse radiation — also vary. Take, for instance, the data cited in Table 33 for several points on USSR territory.

The data in this table confirm the general regularities indicated above, which are expressed in the fact that the annual influx decreases with increasing latitude. Figure 60, which gives the latitude distribution of the total radiation, shows this clearly. Data for individual stations are represented by points and the mean variation of the total radiation with latitude by curve A. For comparison against the actual values of the total radiation, the influx of heat at the top of the atmosphere is indicated by the curve B, which characterizes the latitude distribution of the incoming radiation at the boundary of the atmosphere.

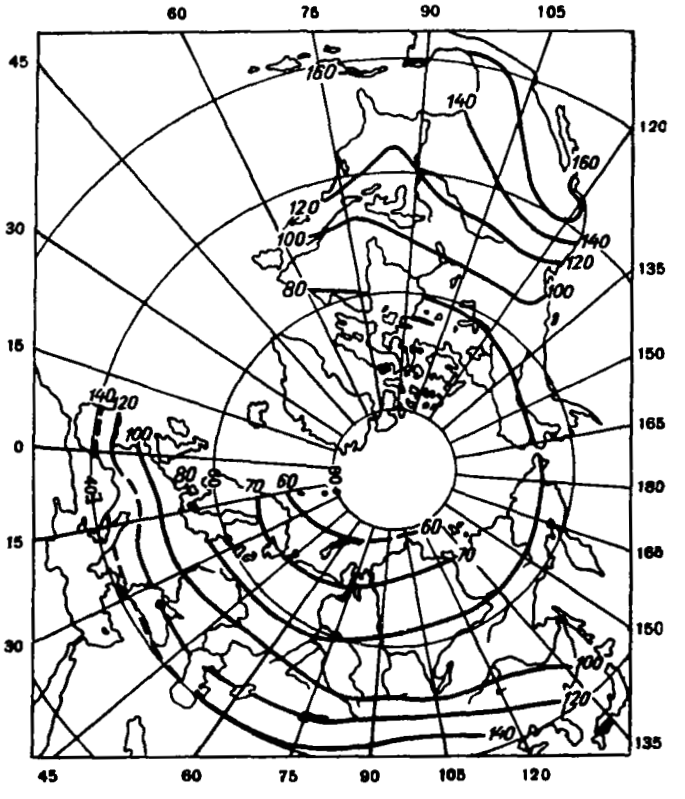


FIGURE 61. Total radiation per annum (kcal/cm^2)

The magnitude of the contribution of diffuse radiation to the over-all influx of solar energy is evident from Table 33. In the north the mean annual influx of diffuse radiation may exceed the influx due to direct radiation by a factor of nearly 2.5 (244%).

If data on the sum of total radiation are available one can map its distribution over a more or less extensive region in individual seasons or as a yearly average.

For illustration we present a map of the annual distribution of total radiation over the extra-tropical land masses of the northern hemisphere (Figure 61). The numbers placed at the end of the isograms indicate the total radiation, in kilocalories per square centimeter of horizontal surface per year. As one can see, total radiation increases from roughly 60 kcal/cm^2· ·year in the far north to 140 kcal/cm^2·year in the south (Mediterranean Sea, Central Asia, Mongolia) and even 160 kcal/cm^2·year in America. On the Eurasian continent isograms run predominantly along the parallels.

Chapter 10

LONG-WAVE RADIATION OF THE EARTH'S SURFACE AND ATMOSPHERE. THE RADIATION BALANCE

In addition to the short-wave radiation considered in the preceding chapter, the atmosphere is traversed by long-wave radiation emitted by both the earth's surface and the atmosphere.

Just as direct and diffuse short-wave radiation is almost the only source of energy for the earth, long-wave emission is the agency by which the earth's surface and atmosphere release heat into outer space. Moreover, long-wave emission is responsible for the exchange of heat between the surface and the atmosphere.

This long-wave "thermal" radiation is the subject of the present chapter.

§ 1. *Radiative properties of different regions of the earth's surface and atmosphere*

Every part of the earth's surface and every volume of air at a temperature different from absolute zero emits thermal radiation.

At the temperatures prevailing at the earth's surface (roughly from 200 to 330° abs.), most (up to 99%) of black-body emission is contained in the range 4–40μ. The infrared emission localized in this region of the spectrum constitutes the long-wave radiation of the earth and atmosphere

The total flux of black-body radiation can be calculated from the formula $E = \delta T^4$ cal/cm^2 · min. This flux reaches values comparable to the flux of direct solar radiation S' on a horizontal surface. Thus, at $T = 273$°K (0°C), $E = 0.453$ cal/cm^2 · min; as the temperature rises to 300° the flux increases to 0.660 cal/cm^2 · min.

However, it should be noted that neither the earth's surface nor the atmosphere can be regarded as perfectly black radiators. Research into the long-wave spectra of the various substances and covers composing the surface shows that the earth's surface may be treated with sufficient accuracy as a "gray body", i.e., one can assume that the long-wave radiation of the surface differs from black-body radiation at the same temperature by a factor which is the same for all wavelengths. Thus the Stefan-Boltzmann formula modified to apply to the earth's surface is as follows:

$$E_S = \delta_S \sigma T^4, \qquad (1)$$

where E_S is the flux of surface radiation and δ_S the relative emissivity power of the earth's surface.

Laboratory investigations and measurements under natural conditions show that δ_S varies roughly from 0.85 to 0.99 for different natural covers and that the value $\delta_S = 0.95$ can be taken as average. The highest emittance, and therefore absorptance ($\delta_S = 0.99$) corresponds to freshly fallen pure snow, which in the long-wave spectral region under consideration can be regarded as a perfectly black radiator.

The question of the emission and absorption of radiation in the atmosphere is far more complex. As we noted in Chapter 8 (§ 3), the absorption spectrum of the atmosphere in the infrared region under consideration is highly complex and is determined by the content and absorptance of water vapor, carbon dioxide, ozone and other (secondary) components of the atmosphere.

A great many theoretical investigations have been devoted to clarifying the general regularities of absorption of radiation in the atmosphere. To characterize the absorption of long-wave radiation by an arbitrary layer of the atmosphere one usually makes use of the absorption function (cf. Chapter 7, § 2):

$$A_0(w) = \frac{I(0) - I(w)}{I(0)} = 1 - \frac{I(w)}{I(0)} \qquad (2)$$

and the related transmission function

$$P_w = \frac{I(w)}{I(0)} = 1 - A(w), \qquad (3)$$

where $I(0)$ is the intensity of the radiation incident on a certain layer, w the mass of the absorbing substance and $I(w)$ the intensity of the emergent radiation from the layer.

Researchers have been seeking to establish the form of the absorption function. However, this is a very difficult problem in view of the complexity of the absorption spectrum, which is composed of a series of bands and lines, many of them superposed on one another. Furthermore, it must be remembered that monochromatic emission (absorption) is practically never observed. Emission always takes place in a certain narrow region of the spectrum forming a spectral line of definite shape, within which the absorption coefficient depends on the wavelength λ. Thus there are objections to using the relation $I(w) = I(0)e^{-kw}$ for determining the absorption over the entire line with the assumption that k is independent of the wavelength. Elsasser has shown that the absorption of radiation in an individual spectral line can be taken to be proportional to the square root of the mass of absorbing substance crossed by the radiation, i.e.,

$$A = c'\sqrt{w}, \qquad (4)$$

where c' is a constant for the given line.

This "square root" law was exploited by many authors to calculate absorption. Elsasser's result can also be generalized to the case of absorption due to a series of nonoverlapping lines. But since real absorption bands in the atmosphere have a far more complex structure and consist of overlapping lines, theoretical calculation of absorption in the atmosphere can be carried out only with the assumption of a certain schematized band structure.

The band model most frequently used for calculations is the one suggested by Elsasser, which consists of a system of equidistant lines of equal

intensity (Figure 62). The absorption function for this schematized spectrum was given by Elsasser in the form $A=\Phi\left(\sqrt{\frac{l w}{2}}\right)$, where Φ is the error function of argument $\sqrt{\frac{l w}{2}}$. In this expression $l=\frac{2\pi\alpha s}{d^2}$, according to Elsasser, is the so-called generalized band absorption coefficient, α being the half-width of the absorption line, d the distance between lines in the schematized band in question, s the total intensity of the line and, finally, w the mass of absorbing substance. This model, however, is highly artificial.

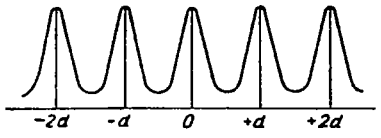


FIGURE 62. Schematized absorption band

Another way of schematizing the structure is based on a statistical model of the absorption band, assuming a random distribution of the intensities and positions of individual lines. This approach was adopted by R. M. Goody, W. L. Godson, L. D. Kaplan and many other foreign scientists. However, the transmission functions obtained as a result of their researches are very complicated. Moreover, they require the determination of many quantities as yet scarcely known.

In view of this simpler methods based on an empirical schematization of absorption in the atmosphere are being used. The function most frequently used to describe the absorption of long-wave radiation in the atmosphere is the exponential absorption function. The entire infrared spectrum is broken down into many (up to 40) narrow spectral regions in each of which the absorption coefficient is assumed to be independent of wavelength. Absorption is described by the exponential function $I=I_0e^{-k_jw}$, where the absorption coefficient k_j for each region is given by $k_j=\frac{1}{w}\ln\frac{I_0}{I}$. If the number of regions is n and an arbitrary j-th region receives a p_j-th part of the flux of radiant energy originally incident on the layer, the integral absorption function can be written as follows:

$$A=1-\sum_{j=1}^{n}p_je^{-k_jw}. \qquad (4')$$

This schematization of the infrared spectrum has been carried out in a variety of ways by different authors with emphasis on the principal absorbing component of the atmosphere, water vapor. The influence of carbon dioxide and ozone is allowed for either separately or by introducing appropriate corrections. While the influence of ozone may be disregarded in view of its smallness (of the order of 2–3%), the same is not true of carbon dioxide.

In the schematization proposed by K. Ya. Kondrat'ev the entire spectrum of water vapor is divided into four sectors for which the following mean absorption coefficients are taken (cm^2/g): $k_1=0.10$, $k_2=1.14$, $k_3=1.96$, $k_4=114$, [sic]; the integral absorption function is given in the following form

$$A=1-\frac{1}{4}\sum_{j=1}^{4}e^{-k_jw}.$$

This simplified absorption function can be employed for all approximate calculations.

Thus the function for absorption of long-wave radiation by water vapor can be expressed in either of two ways, as (4) or (4'); the latter is evidently the more general form.

The foregoing is valid for directed radiation. However, fluxes of radiation in the atmosphere are diffuse, a fact which must be taken into account. Assuming monochromatic radiation, the attenuation of which is expressed by the relation $J_\lambda = J_{0,\lambda}\, e^{-k_\lambda w \sec\vartheta}$, where ϑ is the angle characterizing the direction of propagation of the radiation, the transmission of the flux of diffuse radiation is equivalent to the transmission of direct radiation for the same value of w but for $\sec\vartheta = 1.66$ ($\vartheta \approx 53°$). Then the transmission function for monochromatic diffuse radiation can be expressed in the form $P_\lambda = 1 - A J_\lambda = e^{-1.66 k_\lambda w}$.

Due to the complexity of calculations based on theoretical formulas, in practical work the emission (absorption) of the atmosphere is calculated by means of empirical formulas and the so-called radiation diagrams (see below).

§ 2. Approximate equations of transfer of radiant energy

Radiative transfer of heat usually takes place between bodies separated by an absorbing, scattering and emitting medium. The calculations required for this process are highly complicated in the general case but become somewhat simpler when one considers the transfer of thermal radiation in the atmosphere, since scattering and also the diffuseness of radiation can then be disregarded.

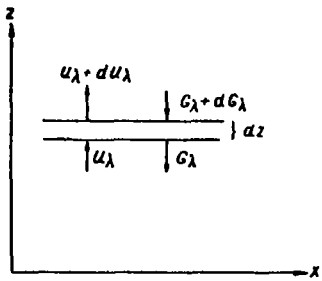


FIGURE 63

Accordingly, let us consider the monochromatic fluxes of long-wave radiation directed upward and downward from a certain level. We denote the descending flux by G_λ and the ascending one by U_λ. Many authors have shown that if only the propagation of radiation in the vertical direction is considered, disregarding the diffuseness of radiation, the latter can be accounted for approximately (as indicated earlier) by multiplying the absorption coefficient by 1.66.

Consider the fluxes of radiation entering and leaving an elementary layer of the atmosphere of thickness dz. These fluxes are shown in Figure 63 (z-axis points upwards). We will first obtain the equation of transfer for the flux G_λ. Owing to absorption in the layer dz the flux G_λ is reduced by the amount $k_\lambda \rho G_\lambda dz$, where ρ is the density of the substance which absorbs the radiation. However, owing to emission in the layer dz the flux G_λ increases by an amount $e_\lambda \rho dz$, where e_λ denotes the spectral emissivity of the layer in question. The net increase is

$$dG_\lambda = k_\lambda \rho G_\lambda dz - e_\lambda \rho dz. \qquad (5)$$

Assuming that Kirchhoff's law holds, i.e., setting $e_\lambda = k_\lambda E_\lambda$ and introducing $dm = \rho dz$, we obtain

$$\frac{dG_\lambda}{dm} = k_\lambda (G_\lambda - E_\lambda). \qquad (6)$$

For the flux U_λ, arguing in a similar way, we find that

$$\frac{dU_\lambda}{dm} = k_\lambda (E_\lambda - U_\lambda). \tag{7}$$

We have thus obtained the approximate equations of transfer of radiant energy for the case of monochromatic long-wave radiation.

Equations (6) and (7) make it possible to calculate the flux of monochromatic radiation in the atmosphere for a specified height distribution of the temperature and density of the absorbing substance (water vapor). From the preceding section it follows that these equations cannot be rewritten so as to apply to the total fluxes $G = \sum_{(\lambda)} G_\lambda$ and $U = \sum_{(\lambda)} U_\lambda$, where summation is over all wavelengths. The total fluxes can be obtained by integrating the given equations over the corresponding region of the spectrum. However, one can simplify the highly complex absorption spectrum of water vapor considerably by dividing it into sectors within which the absorption coefficients vary comparatively little and can therefore be represented by averages. If the schematized absorption band described in the previous section (after Kondrat'ev) is used, then obviously instead of the two equations (6) and (7) one obtains four pairs of equations of the same kind. These have the following form:

$$\frac{dG_j}{dm} = k_j (G_j - p_j E),$$

$$\frac{dU_j}{dm} = k_j (p_j E - U). \tag{8}$$

Here $j = 1, 2, 3, 4$ and p_j is the fraction of black-body emission received by the j-th sector of the spectrum with absorption coefficient k_j. The total fluxes will then be $G = \sum_{j=1} G_j$ and $U = \sum_{j=1}^{4} U_j$.

These equations can be used to solve the problem of the transfer of long-wave radiation in the atmosphere, given the height distribution of temperature and density of the absorbing substance. Assuming a linear variation of the temperature with height and exponential distribution of the density of water vapor, one can obtain formulas for the flux of radiation at any level in the atmosphere.

The unwieldiness of numerical methods of calculation has led to the development of graphical methods. Nomograms, known as radiation diagrams, have been constructed from the equations of transfer. By means of these diagrams it is possible to calculate the fluxes of long-wave radiation and radiative transfer of heat at different levels from the observed distribution of temperature and humidity in the atmosphere.

There are several types of radiation nomograms, the most general being that of A. A. Dmitriev. However, it does not account for the influence of carbon dioxide (the introduction of corresponding corrections is envisaged). In addition the quantitative characteristics of absorption on which the nomogram is based are somewhat out of date.

Currently one of the most refined and convenient nomograms is that of F. N. Shekhter, which makes allowances for absorption and emission by both water vapor and carbon dioxide. The diagrams mentioned here are widely used in the USSR. Abroad the nomograms developed by R. Mügge and F. Möller, and also W. M. Elsasser, are employed.

We note that the accuracy of calculations based on these nomograms is about the same for all and is satisfactory for the solution of most practical problems. This is confirmed by experimental data: recent balloon measurements of the effective emission by V. G. Kastrov and S. S. Gaigerov are in good agreement with data calculated from Shekhter's diagram.

In conclusion, theoretical results obtained by Kondrat'ev for the flux of long-wave radiation at different levels in a cloudless atmosphere are presented in Table 34 (values are in cal/cm^2 · sec · 10^{-4}) [sic]. In calculating these values it was assumed that absolute humidity was 7 g/m^3 at the surface and decreased exponentially with height, and that the vertical temperature gradient in the troposphere (up to 11 km) was constant and amounted to 6°/km with $T = 290°$; F denotes the difference between the fluxes U and G ($F = U - G$).

TABLE 34

Vertical variation of fluxes of long-wave radiation in the atmosphere (cal/cm^2 · min) [sic]

Fluxes	$\frac{z-z_0}{-10^{-4}}$ cm	Height z (cm)					
		$2 \cdot 10^2$	$5 \cdot 10^3$	10^4	10^5	$2 \cdot 10^5$	$1.1 \cdot 10^6$
G	0.483	0.483	0.476	0.472	0.423	0.349	0.021
U	0.574	0.574	0.571	0.569	0.544	0.519	0.392
$F = U - G$	0.091	0.091	0.095	0.097	0.121	0.170	0.371
$\frac{dF}{dz}$	$4.6 \cdot 10^{-7}$	$4.6 \cdot 10^{-7}$	$3.0 \cdot 10^{-7}$	$2.6 \cdot 10^{-7}$	$1.7 \cdot 10^{-7}$	$1.6 \cdot 10^{-7}$	$3.4 \cdot 10^{-8}$

According to the data in Table 34 the total fluxes G and U decrease with height, a natural consequence of the vertical drop in air temperature. On the other hand, the difference between the fluxes U and G, the so-called effective flux F, increases with height nearly throughout the troposphere. Calculations show that, on the average, the total effective flux F increases practically linearly with height throughout the atmosphere.

The last row in the table gives the values of $\frac{dF}{dz}$. This quantity is simply the amount of radiation absorbed by one cubic centimeter of air. In the given scheme of calculation, with a linear decrease of temperature with height, the heat influx is everywhere positive; for some other law of variation of the temperature (e. g., logarithmic) it could be negative in a certain layer.

Fluxes of long-wave radiation in the atmosphere are greatly affected by clouds, particularly continuous low forms. In the presence of clouds there is a marked increase in the descending flux G and, therefore, decrease in the effective flux F. Above the clouds, on the contrary, the transfer of radiation is intensified. Radiative cooling of the upper surfaces of clouds, as we will see later, is responsible for the formation of inversions above clouds. The influence of continuous cloud on long-wave radiation is comparatively simple to estimate quantitatively. In the case of clouds not covering the entire sky and lying at different elevations, the influence on long-wave radiation is far more difficult to compute.

§ 3. Emission of earth's surface and counter-emission of the atmosphere. Effective emission

The difference between the proper emission of the underlying surface and the absorbed part of the counter-emission of the atmosphere is called the effective emission of the underlying surface.

The proper emission of the earth's surface, given by

$$E_S = \delta\sigma T_S^4 , \tag{9}$$

is largely absorbed in the atmosphere. The atmosphere in turn emits long-wave radiation. The total descending flux at the level of the earth's surface is the counter-emission of the atmosphere.

$$E_A = G_{z=0} = G_0. \tag{10}$$

We note that the upward flux of thermal radiation at the top of the atmosphere is called the escaping radiation F_∞.

Denoting the magnitude of the effective emission by F_0, we have

$$F_0 = E_S - \delta E_A. \tag{11}$$

In conformity with the fact that temperatures in the atmosphere are lower than surface temperatures, the fraction of the counter-emission E_A absorbed by the earth's surface exceeds the earth emission E_S only in certain exceptional cases (inversions and clouds). Therefore usually $E_S > E_A$, and the effective emission nearly always characterizes the amount of heat lost by the underlying surface as a result of emission. Thus the effective emission is the difference between two comparatively large quantities: the emission of the earth's surface and counter-emission of the atmosphere.

By now the effective emission has been measured over a period of many tens of years. However, because the number of points at which such measurement is carried out is small, and chiefly because the equipment used is not refined (error of the order of 10–15% in measurements), the material obtained is insufficient and not always completely reliable. Moreover, usually the emission of the black surface of an instrument, and not that of the earth's surface, is measured. Finally, owing to the difficulties involved in removing the short-wave radiation incident on the instrument when measuring the long-wave radiation E_A, the effective emission has been measured until recently only at night. It should be kept in mind, however, that effective emission takes place all day long.

From the observations it was established that effective emission is strongly influenced by changing meteorological conditions. The principal ones are: cloudiness; temperature and absolute humidity of the surface layer of air; and also the temperature difference between the underlying surface and the air.

The simplest relations providing a general idea of the magnitude of F_0 are obtained by observations in clear skies. On clear days and at night, the effective emission varies within 0.10–0.20 cal/cm^2·min at different points of the earth's surface (average); at Pavlovsk, for example, it amounts to about 0.125 cal/cm^2·min.

In its daily march the effective emission conforms to the temperature pattern. It decreases monotonically from sunset to sunrise, next increases toward noon with maximum between 12 and 14 hrs local time, then drops off again toward sunset. Maximum values in daytime reach 0.3–0.4 cal/cm^2·min.

As to the counter-emission of the atmosphere, its values under a cloudless sky most often lie near 0.4–0.5 cal/cm^2·min, reaching to about 0.6–0.7 cal/cm^2·min in a few cases. It also displays a comparatively simple march, characterized by a nighttime decrease ending at dawn and subsequent rise in the pre-noon hours to a maximum some time after midday.

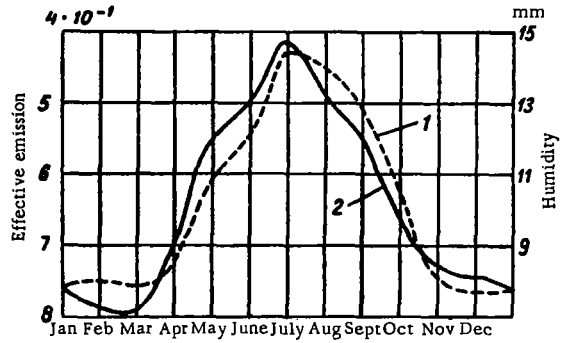


FIGURE 64. Correlation between vapor pressure (1) and effective emission (2)

The annual march of the effective emission is different at different points and is often fairly complicated owing to variations in the total water vapor content of the atmosphere and in the vertical distribution of temperature.

TABLE 35

Clouds	F_0 (cal/cm^2·min)
None	0.147
Cirrus	0.124
Altocumulus	0.047
Low	0.041

Water vapor content and cloudiness have a considerable influence on the magnitude of the effective emission. The influence of the vapor content may be judged from Figure 64, which is based on data of observations at Pavlovsk. It shows that effective emission decreases with increasing vapor pressure. The influence of clouds consists of decreasing the effective emission and increasing the counter-emission. On the average effective

TABLE 36

Cloudiness (scale points)	0	1	2–4	5–6	7	8	9	10
F_0 (cal/cm^2·min)	0.144	0.140	0.132	0.105	0.090	0.083	0.046	0.021

emission under a cloudless sky exceeds that in the presence of cloud by 16–18%, while the counter-emission of a clouded sky exceeds that of a

cloudless sky by 22%. The reason why clouds exert such an influence is that they are a powerful source of radiation. The degree of this influence depends on the amount and type of cloud, as may be seen from Tables 35 and 36.

Empirical formulas for the influence of clouds on effective emission have been suggested by a number of authors (see following section).

The presence of dust, smoke and haze in the atmosphere affects F_0 in the same way as clouds. According to observational data, the associated drop in F_0 can sometimes reach 70–80%. This is widely exploited, in particular, for shielding plants from frost, by filling the air over vulnerable areas with smoke.

§ 4. Methods of calculating the effective emission

Inadequate measurements and complicated theoretical formulas soon led researchers to propose empirical formulas for the calculation of the effective emission. Those suggested by A. Ångström and D. Brunt have gained wide currency. They are based on a close relationship, established by observations, between the magnitude of the effective emission F_0 under clear skies and the atmospheric humidity.

The formula suggested by Ångström for effective emission has the following form for clear skies:

$$F_0 = \sigma T^4 (A + Be^{-Ce}) \tag{12}$$

and for the counter-emission of the atmosphere

$$E_A = \sigma T^4 (A_1 - Be^{-Ce}), \tag{13}$$

where $A_1 = 1 - A$.

In the above formulas σ is an emission constant, T and e are the air temperature and vapor pressure (millimeters) near the surface (at the height of the meteorological screen, 1.5–2.0 m), and A, B and C are empirical coefficients determined by observation or calculation. Different authors use somewhat different values for these coefficients but the ones most frequently used are: $A = 0.194$, $B = 0.236$ and $C = 0.069$. The values suggested by H. M. Bolz and G. Falckenberg are apparently more exact ($A = 0.180$, $B = 0.250$, $C = 0.126$).

The analogous formulas of Brunt are: for effective emission

$$F_0 = \sigma T^4 (a - b\sqrt{e}) \tag{14}$$

for counter-emission

$$E_A = \sigma T^4 (a_1 + b\sqrt{e}), \tag{15}$$

where $a_1 = 1 - a$; the values of the coefficients a and b, which are established empirically, are also fairly variable (0.34–0.66 for a and 0.33–0.127 for b).

Since these formulas do not consider the departure of the earth's surface from a black body and the considerable effect of clouds, the suggestion was made that in (13) the correction factors $(1 - cn)$ and δ be introduced as follows:

$$F_0 = \delta\sigma T^4 (A + Be^{-Ce})(1 - cn). \tag{16}$$

The value of the factor c depends not only on total cloudiness but on cloud form and height as well. Therefore when calculating the effective emission one should make different allowances for clouds at different levels. This was done by many authors by replacing the factor $(1 - cn)$ in formula (16) by the more complex factor

$$[1 - (c_l n_l + c_m n_m + c_h n_h)],$$

where the subscripts "l", "m" and "h" indicate that the corresponding quantities apply to low, middle and high clouds respectively. Most often, however, the coefficient c is taken to be 0.76 from an estimate of the mean cloudiness.

Formula (16) can be made somewhat more exact by simple transformations. Then

$$F_0 = \delta [\sigma T^4 (A + Be^{-ce}) + \sigma (T_0^4 - T^4)] (1 - 0.76n), \qquad (17)$$

where T_0 is the temperature of the underlying surface.

In this form the formula takes into account the influence on effective emission of extremely strong changes ("jumps") of temperature in the layer extending from the surface to the level at which the air temperature T was measured. However, owing to its statistical character even formula (17) is applicable only under certain average conditions.

For greater accuracy than can be attained with empirical formulas the effective emission can be calculated (if data of aerological soundings are available) with the aid of the radiation nomograms mentioned earlier.

A knowledge of the distribution of intensity of effective emission and counter-emission in various directions is important for the solution of a number of practical problems. This distribution can also be found by calculations using radiation nomograms or empirical formulas. One such formula, which was long in use, is that of Linke:

$$f_\vartheta = f_0 \cos^r \vartheta; \qquad (18)$$

here f_ϑ and f_0 are the intensity of effective emission in the direction of the zenith angle ϑ and the zenith respectively, and the index r depends on the vapor pressure at the earth's surface. According to Linke, $r = 0.3$ on the average (this corresponds to $e = 5.4$ mm).

Exploring this question theoretically, Kondrat'ev obtained an expression relating the intensity of effective emission f_ϑ in the direction of the zenith angle ϑ to that in the direction of the zenith f_0. For approximate calculations this relation can be written in the form

$$f_\vartheta = f_0 e^{-kw_\infty (\sec \vartheta - 1)}, \qquad (19)$$

where $k = 0.144$ is a numerical factor representing the mean transmission coefficient and w_∞ is the mass of water vapor in a vertical column of air of unit cross section (1 cm^2), expressed in grams.

The observations, while admittedly still very scanty, confirm the directional dependence of effective emission.

At the same time as it takes on heat by absorption of direct and diffuse radiation, the earth's surface releases it by effective emission. Table 37 gives data on the annual march of the daily sum of effective emission for cloudless skies (cal/cm^2 · day) for three points in the USSR; Table 38

presents data on the annual march of the mean monthly sum of effective emission (kcal/cm^2· month) for clear skies ΣF_0 and average cloudiness ΣF_{cl}.

TABLE 37

Annual march of daily sum of effective emission for cloudless sky (cal/cm^2· day)

Point	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
Yakutsk	180	158	181	180	207	192	190	160	160	192	128	122
Pavlovsk	144	148	180	219	209	202	202	184	171	168	174	173
Karadag (Crimea)	235	233	242	230	204	202	222	209	219	252	248	228

One can see from the tables that the sums of effective emission are everywhere (with some departures) significantly greater under clear skies than for cloudy skies; in both cases sums generally increase southwards.

TABLE 38

Annual march of mean monthly sum of effective emission (kcal/cm^2·month)

	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Year
						Yakutsk							
ΣF_0	5.6	4.4	5.6	5.4	6.4	5.8	5.9	5.0	4.8	6.0	3.8	3.8	61.6
ΣF_{cl}	4.2	3.7	4.6	3.9	3.8	3.9	3.5	3.3	3.4	3.3	2.7	2.7	43.0
						Pavlovsk							
ΣF_0	4.5	4.1	5.6	6.8	6.7	6.1	6.5	5.9	5.2	5.2	5.2	5.0	66.8
ΣF_{cl}	2.0	1.8	3.2	3.8	4.0	3.9	4.2	3.5	2.9	2.4	2.0	2.0	35.7
						Karadag							
ΣF_0	6.3	6.5	7.5	6.9	6.3	6.1	6.9	6.5	6.6	7.8	7.4	7.1	82.9
ΣF_{cl}	3.7	3.7	5.0	4.6	4.9	4.9	6.0	5.3	5.3	5.1	4.2	3.9	56.6
						Tashkent							
ΣF_0	4.7	5.2	6.2	6.4	6.8	7.3	7.0	7.1	6.9	6.3	5.8	5.3	75.1
ΣF_{cl}	3.2	3.4	4.0	4.5	4.7	6.1	6.4	6.8	6.5	5.4	4.5	3.4	58.9

The numerical monthly sums are perceptibly different at different points, and their annual march is fairly complicated. Calculations show that the geographic variations of the mean sum of effective emission are rather poorly defined. The largest annual sums are observed in tropical deserts. Near the equator proper, over both sea and land, effective emission is weaker (about 30 kcal/cm^2 · annum); it increases with increasing latitude to 40–50 kcal/cm^2· annum near lat. 60°, then drops off again.

§ 5. Radiation balance

Radiation balance is the term applied to the income-expenditure of radiant energy absorbed and emitted by the earth's surface, atmosphere or

earth's surface-atmosphere system over different periods of time (minute, day, month, year).

Let us first consider the radiation balance R of the earth's surface. We mention right away that what is really studied here is the radiation balance of a certain active layer (not surface) varying in thickness over a fairly broad range.

The incoming part of the radiation balance R consists of the direct and diffuse solar radiation and atmospheric counter-emission absorbed by the surface. The outgoing part consists of the proper thermal emission of the earth's surface. Denoting the albedo of the surface by A and its absorptivity by δ, we can write the equation of the radiation balance in the form

$$R=(Q+q)(1-A)+\delta E_A-E_S, \tag{20}$$

where Q and q are the fluxes (or sums) of incoming direct and scattered radiation, and E_A and E_S the fluxes (or sums) of atmospheric counter-emission and proper earth emission.

Since $E_S-\delta E_A=F_0$ is the effective emission (flux or sum) of the underlying surface, we have

$$R=(Q+q)(1-A)-F_0. \tag{21}$$

The radiation balance can be determined either selectively by observation of its individual components, or with the help of so-called balance meters; computational methods are also used to find R.

From our earlier discussion we know that the individual components of the radiation balance depend on a number of factors. Hence the balance is also subject to strong variation. It is particularly affected by the character and state of the surface, atmospheric turbidity, water vapor content of the atmosphere, cloudiness and so on.

The radiation balance can be positive (if the influx of heat is greater than the expenditure) or negative (opposite case). In the daily march it is usually positive in daytime and negative at night.

Results of measurements of the daily march of the radiation balance and all its components are shown for illustration in Figure 65. The data were obtained on a clear day (6 September 1945) in an arid spot near Tashkent.

As can be seen from the figure, the variation of the radiation balance at different times of the day is determined primarily by components experiencing maximum fluctuations at these times. By day the variation of the radiation balance is roughly parallel to the variation of direct solar radiation. At night it coincides with the march of effective emission; as the latter changes only slightly, the radiation balance remains nearly constant. Both in daytime and at night, of course, the daily march of the radiation balance is subject to considerable changes under the influence of cloudiness. As a rule the appearance of clouds always causes a drop in the absolute values of the radiation balance; in summer there is a pronounced decrease in the positive daily radiation balance. In winter clouds produce a decrease in the negative daily balance.

The transition from negative nighttime balance to positive daytime balance and vice versa takes place at solar elevations of the order of 10–15° under clear skies and at lower elevations (even lower than 5°) under cloudy skies. The daily radiation balance varies over a very broad range, from large negative values in winter to large positive values in summer.

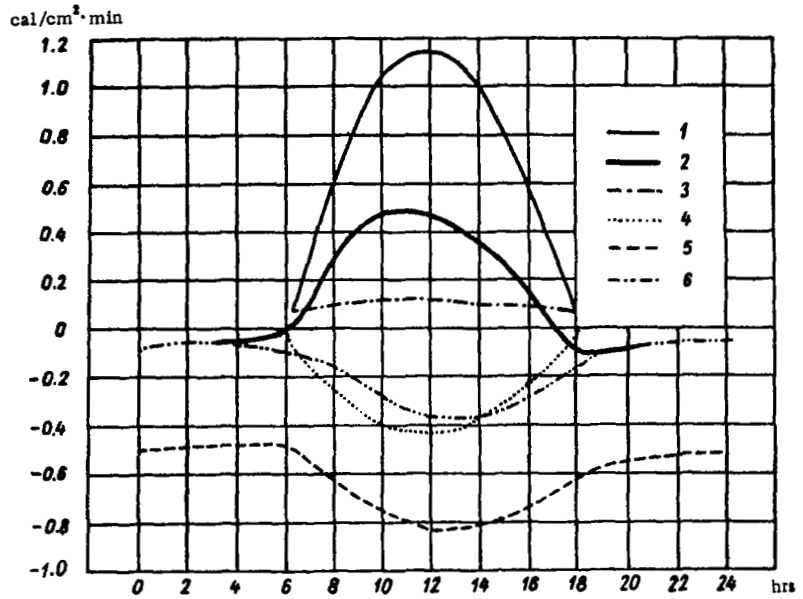


FIGURE 65. Daily march of the components of the surface radiation balance, Arys', 6 September 1945

1–direct radiation; 2–radiation balance; 3–diffuse radiation; 4–reflected radiation; 5–surface emission; 6–effective emission.

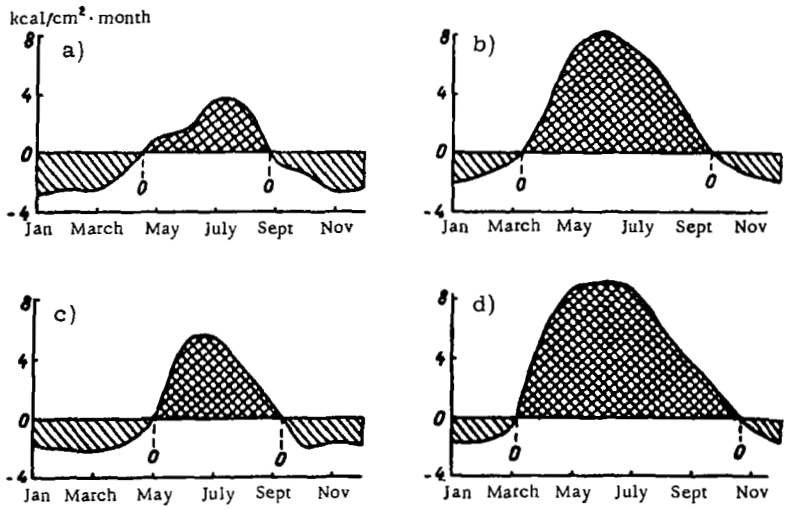


FIGURE 66. Annual march of radiation balance

a–Thikhaya Bay, φ = 80°; b–Tiksi Bay, φ = 72°; c–Leningrad, φ = 60°; d–Saratov, φ = 51°.

The radiation balance of snow covers usually changes fairly little through the day. As a rule it is negative when skies are clear due to the large albedo of snow and strong effective emission of snow covers. The influence of irrigation is also interesting; due to the attendant reduction in the albedo and temperature of the surface, irrigation increases the radiation balance considerably (by 60% in some cases).

The annual march of the radiation balance is illustrated in Figure 66 for four points situated at different latitudes. The radiation balance is everywhere positive in summer and negative in winter, but the annual march is somewhat asymmetric with respect to summer and the annual amplitude increases southward.

Negative values of the balance persist for longer periods in the north. At Kikhaya Bay, for instance, a positive balance is recorded only in four summer months; negative balances persist only for three months at Saratov and still less at Tashkent.

The geographic distribution of the annual sums of the radiation balance can be judged from Table 39, compiled by M. I. Budyko on the basis of calculations.

TABLE 39

Annual sums of radiation balance at earth's surface (kcal/cm^2·annum) after M.I. Budyko

Latitude	Surface radiation balance			Latitude	Surface radiation balance		
	sea	land	entire hemisphere		sea	land	entire hemisphere
60–50° N	34	23	28	10–20° S	107	69	99
50–40	54	38	46	20–30	94	62	87
40–30	78	56	69	30–40	73	55	71
30–20	100	64	86	40–50	53	39	53
20–10	110	74	101	50–60	31	26	31
10–0	107	79	101	Entire earth	77	46	68
0–10° S	107	75	99				

As this table and more detailed study of the problem show, annual sums increase regularly with decreasing latitude. The distribution is approximately zonal over the oceans and somewhat more complicated over land. Maximum values for the earth's surface are observed north of the Arabian Sea (> 140 kcal/cm^2·annum); on land maximum values are observed in the humid tropics (about 100 kcal/cm^2·annum).

In the European USSR the annual sum of the radiation balance varies on the average from 40–45 kcal/cm^2·annum in the south to 5–10 kcal/cm^2·annum in the north.

This south-north change takes place fairly evenly with the exception of the region near the Arctic Circle, where a very rapid northward drop is observed.

Let us now consider the radiation balance of the atmosphere and earth-atmosphere system.

The incoming part of the radiation balance R_A of the atmosphere consists of the direct and diffuse radiation q' absorbed by a vertical column of atmosphere of unit cross section, and the thermal emission of the underlying surface U_T absorbed by the atmosphere. The outgoing part is determined by losses due to thermal emission by the atmosphere in the direction

of the earth's surface δE_A and outer space U_∞. Thus

$$R_A = U_T + q' - \delta E_A - U_\infty. \quad (22)$$

Denoting the transmissivity of the atmosphere to thermal radiation by P, one can write $U_T = (1-P)E_S$, and therefore

$$R_A = E_S - PE_S + q' - \delta E_A - U_\infty. \quad (23)$$

But $E_S - \delta E_A = F_0$ is the effective emission of the underlying surface, and $PE_S + U_{c\infty} = F_\infty$ gives the atmospheric and ground emission escaping into outer space. Taking this into account we have

$$R_A = F_0 - F_\infty + q'. \quad (24)$$

The radiation balance of the earth-atmosphere system R_S can be expressed as the sum of the radiation balance of the surface R and radiation balance of the atmosphere R_A. Bearing in mind (21) and (24) we obtain

$$R_S = R + R_A = (Q+q)(1-A) + q' - F_\infty. \quad (25)$$

The incoming part of the radiation balance consists of the direct and diffuse radiation absorbed by the surface $(Q+q)(1-A)$ and atmosphere q' and the outgoing part of the escaping emission F_∞. It should be mentioned that the radiation balance of the atmosphere and earth-atmosphere system cannot be measured directly and are determined only by calculation.

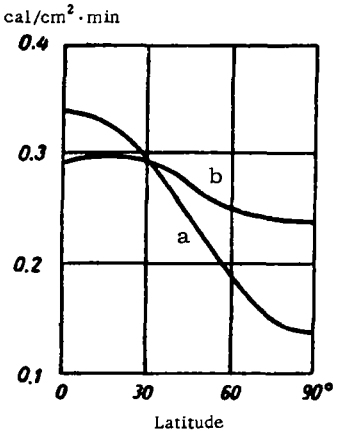


FIGURE 67. Variation of absorbed solar (a) and escaping long-wave (b) radiation with latitude

The radiation balance of the earth's surface-atmosphere system may be positive or negative at individual points. However, since marked deviations of the thermal regime of the earth as a whole from the stationary state are absent, it must be zero for the earth as a whole. Hence since the radiation balance of the earth's surface is positive (68 kcal/cm² · annum), the radiation balance of the atmosphere is the same but negative.

Figure 67 illustrates the latitudinal variation of the absorption of solar radiation (curve a) and escaping radiation (curve b). The radiation balance

of the earth-atmosphere system, as one can see, is positive in the northern hemisphere between the equator and (roughly) lat. 30°. North of this it becomes negative. This distribution should obviously result in a transfer of heat from equator to pole. By way of supplement to Figure 67 results of approximate calculations of the latitude distribution of escaping radiation (cal/cm^2·min) are given in Table 40.

TABLE 40

Latitude distribution of incoming and escaping radiation (cal/cm^2·min)

φ°	Incoming radiation	Escaping radiation	φ°	Incoming radiation	Escaping radiation
0	0,339	0,271	50	0,232	0,277
10	0,334	0,282	60	0,193	0,272
20	0,320	0,284	70	0,160	0,260
30	0,297	0,284	80	0,144	0,252
40	0,267	0,282	90	0,140	0,252

Notwithstanding the tentative nature of calculations of R_S we present the most complete data characterizing the latitudinal march of the daily sum $(Q+q)(1-A)$ in Table 41; the values given are annual averages (cal/cm^2· day). According to N. A. Bagrov, the radiation balance for the entire earth amounts to 0.95 cal/cm^2· day. It should be mentioned that in addition to its variation with latitude R_S also displays a dependence on longitude. This dependence, however, is complex and relatively unstudied.

TABLE 41

Latitude distribution of R_S (cal/cm^2·day), after N. A. Bagrov

Northern hemisphere				Southern hemisphere			
latitude	$(Q+q)(1-A)$	F_{∞}	R_S	latitude	$(Q+q)(1-A)$	F_{∞}	R_S
90–80	156	356	–200	0–10	518	425	93
80–70	184	357	–183	10–20	528	432	96
70–60	240	388	–148	20–30	491	430	61
60–50	294	403	–109	30–40	428	419	9
50–40	358	416	–58	40–50	361	410	–49
40–30	435	432	–3	50–60	278	396	–118
30–20	491	440	51	60–70	214	380	–166
20–10	518	438	80	70–80	171	266	–195
10–0	518	426	92	80–90	156	358	–202

The overall scheme of the mean annual radiation balance will be given later in connection with the thermal balance of the earth, of which it is the principal component.

Part Three

THERMAL REGIME OF THE ATMOSPHERE AND EARTH'S SURFACE

The greater part of solar radiation, as we saw earlier, is absorbed by the surface, which is itself a source of thermal radiation. The atmosphere (if one excludes the uppermost layers) is heated only to a limited extent by the absorption of solar radiation, and its main source of heat is the surface. Heat transfer is constantly taking place between the earth's surface and atmosphere, as well as between individual layers of the atmosphere. Part Three deals with the main features of heat transfer in the atmosphere, and also with the thermal regime of the atmosphere and upper layers of the earth's surface.

Chapter 11

HEAT TRANSFER IN THE UPPERMOST LAYERS OF SOIL AND WATER. THERMAL REGIME OF SOIL AND WATER

The thermal regime of the earth's surface is chiefly a function of the income and expenditure of radiant energy, which is expressed by the equation of radiation balance and depends on the amount of incident radiation, albedo and effective emission. The heat absorbed by the earth's surface is transmitted to adjoining layers of the atmosphere and underlying layers of soil and water and determines the temperature regime of these layers. Data on the temperature of the upper layers of the soil are very important for estimating the temperature of the surface layer of air directly in contact with the soil surface. Such data are also of great practical importance for construction and especially agriculture, since the root systems of all plants lie in the upper layers of soil. Accordingly the question of soil temperature is discussed in detail in textbooks of agricultural meteorology and climatology. In the present chapter we will consider only the main features of heat propagation in the soil.

§ 1. Propagation of heat in the soil

The principal mechanism by which heat is transferred in the soil is molecular conduction.

In order to clarify the main aspects of heat transfer in the soil we will assume that the uppermost layers of the soil are homogeneous and isotropic and that temperature varies only with depth (remains constant in the horizontal direction). We single out a vertical cylinder of unit cross section in the soil layer under consideration (Figure 68). The top of this cylinder represents an elementary area of the earth's surface. We direct the z axis downwards. Then at the depth z the flux of heat B in the soil across the section CD (i.e., the amount of heat moving across a unit surface per unit time along the normal to it) at a certain time will be proportional to the vertical temperature gradient of the soil $\frac{\partial \vartheta(zt)}{\partial z}$ and will be expressed in the form

$$B = -\lambda \frac{\partial \vartheta (zt)}{\partial z}, \qquad (1)$$

where λ is the coefficient of calorimetric conductivity of the soil, which, as is known, is numerically equal to the amount of heat passing through one square centimeter of surface per second for a vertical temperature gradient of 1°/cm (i.e., λ has the dimensionality cal/cm · sec · deg).

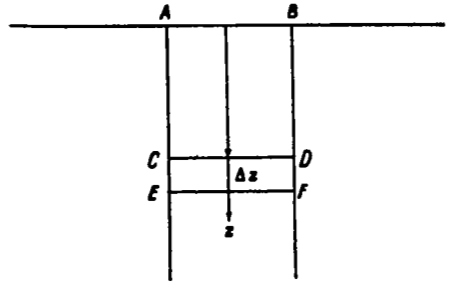


FIGURE 68

The influx of heat to a unit volume (1 cm^3) of soil per unit time is equal to the increment of the flux with opposite sign, i.e., $-\frac{\partial B}{\partial z}$: in view of (1) this can be written as

$$\frac{\partial B}{\partial z} = \frac{\partial}{\partial z}\left(\lambda \frac{\partial \vartheta}{\partial z}\right).$$

The influx of heat can also be expressed as the change in the amount of heat in a unit volume of soil per unit time, i.e., $c\rho \frac{\partial \vartheta}{\partial t}$, where c is the specific heat (the amount of heat necessary to heat one cubic centimeter of soil by one degree centigrade) and ρ the soil density. The volumetric heat capacity $c_{vol} = c\rho$ is often employed.

Bearing the above in mind we obtain $\frac{\partial}{\partial z}\left(\lambda \frac{\partial \vartheta}{\partial z}\right) = c\rho \frac{\partial \vartheta}{\partial t}$, whence

$$\frac{\partial \vartheta}{\partial t} = \frac{\partial}{\partial z}\left(\frac{\lambda}{c\rho} \frac{\partial \vartheta}{\partial z}\right) \approx \frac{\partial}{\partial z}\left(a \frac{\partial \vartheta}{\partial z}\right), \tag{2}$$

where $a = \frac{\lambda}{c\rho} = \frac{\lambda}{c_{vol}}$ is the coefficient of thermometric conductivity.

If λ and c_{vol} do not change with depth the last equation can be written as

$$\frac{\partial \vartheta}{\partial t} = a \frac{\partial^2 \vartheta}{\partial z^2}. \tag{2'}$$

Equation (2) is a particular form of the general equation of heat conduction. In the simplest case, which was studied by Fourier, the parameters c, ρ and λ can be regarded as constant and the equation has the form (2'). The general solution of this equation (Fourier equation) may be found in textbooks of mathematical physics.

Let us assume that in the first approximation the temperature at the soil surface varies periodically in time according to the law expressed by the simple harmonic function

$$\vartheta(0, t) = \vartheta_0 + A_0 \sin \frac{2\pi}{T} t, \tag{3}$$

where ϑ_0 is the mean (daily or annual) value of the temperature at the soil surface, T the period of oscillation (day or year), A_0 the amplitude of the wave and t the time.

Further, let us assume that at a certain great depth (mathematically for $z=\infty$) the daily or annual temperature wave dies out. For these boundary conditions we obtain the solution

$$\vartheta(z, t)=\vartheta_0+A_0 e^{-z\sqrt{\frac{\pi}{aT}}}\sin\left[\frac{2\pi}{T}\left(t-\frac{z}{2}\sqrt{\frac{T}{a\pi}}\right)\right]. \qquad (4)$$

One can show that the solution is correct by introducing (4) into (2) and into the boundary conditions (3).

Analysis of (4) leads to the following conclusions.

a) The temperature at various depths in the soil varies periodically with the same period T as the temperature wave at the soil surface; however, the amplitude decreases with depth according to the law

$$A_z=A_0 e^{-z\sqrt{\frac{\pi}{aT}}}$$

or

$$\ln A_z=\ln A_0=-z\sqrt{\frac{\pi}{aT}}, \qquad (5)$$

i.e., exponentially. Further, when the depths z increase in arithmetic progression $(z, 2z, 3z, \ldots)$ the amplitude decreases in geometric progression $(A_0 : A_z : A_{2z} : \ldots = \lambda : e^{-z} : e^{-2z} : \ldots)$.

In different media (characterized by different values of the thermometric conductivity a) the depths at which the amplitude of a certain wave decreases (in comparison with the amplitude at the surface) by the same factor is given by

$$z_1\sqrt{\frac{\pi}{a_1 T}}=z_2\sqrt{\frac{\pi}{a_2 T}},$$

whence

$$z_1 : z_2=\sqrt{a_1} : \sqrt{a_2}, \qquad (6)$$

i.e., the ratio of the depths at which the same temperature wave is equally damped in media with different thermometric conductivities is equal to the ratio of the square roots of the coefficients of thermometric conductivity $\sqrt{a}$.

b) For oscillations having different periods T_1 and T_2, equal damping will take place at the depths z_1 and z_2, which are related by $e^{-z_1\sqrt{\frac{\pi}{aT_1}}}=e^{-z_2\sqrt{\frac{\pi}{aT_2}}}$, whence

$$z_1 : z_2=\sqrt{T_1} : \sqrt{T_2}, \qquad (7)$$

i.e., the ratio of the depths at which the amplitudes of temperature waves of different periods constitute the same fraction of the amplitude at the surface is equal to the ratio of the square roots of the periods. For example, annual oscillations with a period of 365 days would be damped to a value equal to the damping of daily oscillations (period of 1 day) at a depth $z_{year} : z_{day} = \sqrt{365 : 1} = 19.1$ times greater.

c) The phase characterizing the retardation of the oscillations with

depth is expressed by the relation

$$\varphi_z = z\sqrt{\frac{\pi}{aT}} \tag{8}$$

or, in time units,

$$\varphi_z = \frac{z}{2}\sqrt{\frac{T}{a\pi}}. \tag{8'}$$

The time of occurrence of the temperature maximum at any depth $t_z = t + \frac{z}{2}\sqrt{\frac{T}{a\pi}}$, hence the retardation of the instants of occurrence of temperature extrema (maximum or minimum), given by $t_z - t = \frac{z}{2}\sqrt{\frac{T}{a\pi}}$, is proportional to the depth z and therefore the following relation holds for the two depths z_1 and z_2

$$(t_1 - t) : (t_2 - t) = z_1 : z_2, \tag{9}$$

or, setting $t=0$,

$$t_1 : t_2 = z_1 : z_2. \tag{9'}$$

The above elementary theory of heat propagation in the soil is correct only with regard to the principal features of this process, which is far more complex under actual conditions. In fact periodic variations of the temperature at the soil surface cannot be described exactly by a simple sinusoid, as we have done above; moreover, they are subject to random variations. For greater accuracy, of course, one could replace the simple boundary condition (4) by a more complex one, expressing it in the form of a harmonic series. This substitution would lead to a more complicated form of solution in which the formulas obtained for the amplitude and phase would be analogous to those derived above. However, in these formulas, in addition to oscillations with fundamental period T, one would also have its harmonics with periods $\frac{T}{2}$, $\frac{T}{3}$, etc. It should also be borne in mind that molecular conduction is not the only mechanism of heat transfer in the soil. However, the role of the other possible mechanisms of heat transfer (convective and radiative) is generally so small that it can be disregarded in practice. But the most important point which must always be remembered when using the above results is that the coefficients c, λ and a characterizing the thermal properties of the soil are not, as we have assumed, constant, but vary in a complex manner both with depth and with time. This is very difficult to allow for theoretically. Consequently, the problem of heat transfer can be solved with sufficient accuracy only for individual particular cases, for a specified law of variation with depth and time of the parameters characterizing the thermal properties of the soil.

§ 2. Thermal properties of the soil

From the discussion in the preceding section it is obvious that any investigation into the thermal regime of the soil requires accurate knowledge of its thermal properties, given by the value of its volumetric heat capacity $c_{vol} = c\rho$ and calorimetric conductivity λ, or by the value of the thermometric

conductivity a. However, since experimental determination of these parameters under natural conditions, which are highly variable in time, is a complicated matter from the methodological standpoint, only a few such measurements have been performed. A somewhat larger volume of data is available for the thermal properties of individual soil samples under laboratory conditions, but their practical value is severely limited.

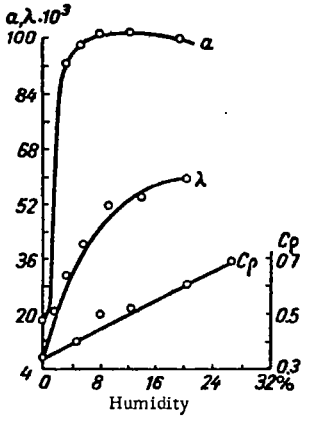


FIGURE 69. Humidity dependence of thermal characteristics of soil

The difficulty of the problem arises from the fact that the thermal properties of a soil depend not only on its chemical-mineralogical composition and mechanical structure, but also largely on the percentage of pore space occupied by water and air.

Investigation into the heat capacity of the various mineral components of the hard skeleton of the soil have shown that this quantity varies only slightly; on the average it amounts to about 0.19 cal/g · deg. The variation of the calorimetric conductivity is also small, although its range is somewhat broader. Its mean value is roughly 10^{-3} cal/cm · sec · deg. If we recall for comparison that the volumetric heat capacity of water is about unity and that of air three to four thousand times smaller, and that the calorimetric conductivity of water is about $130 \cdot 10^{-5}$ cal/cm · sec · deg and that of air about $5-6 \cdot 10^{-5}$ cal/cm · sec · deg, it becomes obvious that the thermal properties of a soil will depend not so much on its mineralogical composition as on its porosity, which is characterized by the ratio of the pore volume to the volume of the soil sample as a whole (expressed in percents), and on the fraction of pore space occupied by water or air.

From the data mentioned above it is evident that both the heat capacity and the calorimetric conductivity of different kinds of soil always increase with increasing moistness of the soil. The variation of c_{vol} and λ with varying soil humidity is illustrated by Figure 69, which gives data for fine loam. One can see that the heat capacity increases continuously, though comparatively slowly, with increasing humidity. The curve of the calorimetric conductivity as a function of humidity is damped with rapid growth at the beginning of moistening and slower growth later, and at the limit the calorimetric conductivity of soil approaches that of water. The peculiar behavior of the coefficient of thermometric conductivity is due to the joint influence of the variation of the heat capacity and calorimetric conductivity.

The calorimetric conductivity λ decreases as soil porosity increases; this is due to the increased content of air in the soil. Despite the considerable differences in mineralogical composition between different soils, it was found that in solving practical problems one could restrict oneself to considering only three categories of soil types according to degree of dispersion: finely dispersed— clays (poor conductors of heat), medium dispersed— sandy loams and loams, and coarsely dispersed— sands (good conductors). When considering the thermal properties of soil it is necessary to take into account the influence of porosity and especially humidity, which in certain kinds of soil may sometimes alter its thermal characteristics by 100%.

We give representative values of the thermal characteristics for soils subjected to varying degrees of moistening in Table 42. These values can be used for approximate calculations.

TABLE 42

Representative values of λ, c, ρ and a for soil with different degrees of moistening

Degree of soil moistening	$c_{vol} = c\rho$ (cal/cm^3 · deg)	λ (cal/cm · sec · deg)	$a = \frac{\lambda}{c\rho}$
Dry	0.32	0.0005	0.0016
Slightly moist	0.38	0.0011	0.0020
Fairly moist	0.46	0.0020	0.0043
Very moist	0.58	0.0035	0.0060

We note that the temperature of a soil can have a certain influence on its calorimetric conductivity; however, this is negligible in comparison with the influence of humidity and porosity.

Since soil humidity and the degree of compactness of the soil (i.e., ρ) change with depth, the volumetric heat capacity $c\rho$ and calorimetric conductivity λ also change with depth. The nature of this variation can be very different in different cases.

§ 3. *Heat transfer in the soil*

Heat transfer takes place between the earth's surface and underlying soil layers at any instant. This is quantitatively characterized by the flux of heat (cf. § 1) which flows either from the surface to the interior of the soil or in the opposite direction depending on the sign of the vertical temperature gradient. The flux directed from the surface into the soil is regarded as positive.

A positive heat flux is obviously related to the decrease of soil temperature with depth. The corresponding temperature distribution in the soil is called the insolation type. It is characteristic of instants and periods of time at which the absorption of radiation by the earth's surface is greater than the emission (usually in daytime and in summer). Negative heat transfer in the soil, on the contrary, is related to an increase in soil temperature with depth. Such a temperature distribution in the soil is called the emission type. It is characteristic of instants at which the emission of the surface is greater than its absorption (winter, night). Owing to damping of temperature oscillations (daily and annual) with depth, the process of heat transfer takes place in a certain soil layer which increases in thickness with the period of oscillation. The layer in which annual and daily temperature oscillations take place is usually called the active layer of the soil. The temperature at its lower boundary remains constant in time, and the so-called layer of constant annual temperature lies there. Below the layer of constant annual temperature begins a continuous rise in temperature towards the central regions of the earth, characterized by the geothermal temperature gradient; this gradient amounts to about 0.0003°/cm in bed rock. The total

amount of heat delivered by the earth's interior to the soil surface is very small (about 55 cal/cm^2 · annum) compared with solar radiation.

The flux of heat in the soil $B = -\lambda \frac{\partial t}{\partial z}$ is determined by the thermal properties and temperature distribution in the soil. It is usually calculated from the values of these quantities, which are determined by direct measurement or by calculation on the basis of very simple meteorological observations.

At stations of the USSR meteorological network formulas for the calculation of the flux B from measurements of soil temperature and humidity are employed for practical computations (these formulas may be found in the handbooks). Observational data lead to the following conclusions.

In absolute magnitude both positive and negative values of B amount to several hundred cal/cm^2 · min. In the daily march the flux of heat is positive in daytime and negative at night, and extremum values occur at different hours in different months. Maxima are recorded in the pre-noon hours (8–11 hrs) and minima in the late afternoon (19–22 hrs), shifting to earlier and later hours respectively as summer approaches.

The most characteristic features of the daily march of heat transfer in the soil are: very considerable variations in daytime and a nearly constant transport at night. The reasons for this lie in the character of the radiation balance of the soil surface. In daytime the flux of heat toward the soil varies considerably with time, chiefly owing to the variation of the influx of solar radiation. At night the thermal flux in the soil changes comparatively little.

TABLE 43

Annual march of the mean daily heat flux

Months	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
Surface temperature (°)	-1.5	0.9	5.5	11.5	17.1	20.9	21.7	19.4	14.5	8.5	2.9	-0.7
Heat flux toward soil (kcal/cm^2 · day)	-7.3	-4.4	2.5	12.9	18.3	14.6	7.3	1.3	-5.5	-12.9	-15.6	-12.1

The values of the heat transfer cross zero (from negative at night to positive in daytime and back) at times corresponding closely to sunrise and sunset. In the morning transition occurs some time after sunrise and in the evening 1–2 hours before sunset. The magnitude of the heat flux is considerably greater on clear days than on overcast days.

TABLE 44

Months	Jan-Feb	Feb-March	March-April	April-May	May-June	June-July	July-Aug	Aug-Sept	Sept-Oct	Oct-Nov	Nov-Dec	Dec-Jan
(cal/cm^2 · month)	-300	-160	-6	353	498	469	345	147	-133	-386	-425	-394

Considering the sum total over a day or month, the soil accumulates heat in the warm season and releases it in the cold season. This is shown by

Tables 43 and 44, which give data for Dresden (daily mean for each month of the year) and a sandy soil in middle latitudes (monthly values) respectively.

From these data one can see that the maximum positive heat transfer takes place in May–June and the maximum negative heat transfer in November–December.

We note that in middle latitudes the period of accumulation of heat in the soil is shorter (5 months) than the period of expenditure (7 months).

The total amount of heat transported annually across the soil is considerable. From Table 44 it is seen to be about 1810 cal/cm^2· annum. Observations at Voeikovo (Leningrad area) give values of the same order. Computations by G. A. Lyuboslavskii for Lesnoi (Leningrad) gave an annual influx of 1471 cal/cm^2 and expenditure of 1412 cal/cm^2; according to the latter, the difference (59 cal) was removed by ground water.

The annual march of heat transfer in the soil is also affected to a large extent by the snow cover, a factor we will discuss below.

§ 4. Daily and annual march of soil temperature

We now turn to results of the direct observation of temperature at various depths in the soil.

Many of the measurements considered were carried out on a bare surface, i.e., from which vegetation, and in winter also snow, had been artificially removed.

Once again, when we use the term soil surface we always mean the so-called "active surface", a certain surface layer of soil of variable thickness which directly absorbs solar and atmospheric radiation and releases heat by emission into the atmosphere. Measuring the temperature of this active layer is an extremely difficult task even in the simplest case, where the active surface consists of a smooth, plant-free bare soil. In principle the most refined way of determining the temperature of the active surface is by the radiometric method, in which the temperature of the surface under consideration is evaluated from the amount of radiation emitted by it.

In view of the difficulties involved in measurement, all data obtained so far on the temperature of the soil surface are to be regarded as approximate.

In the daily march the maximum temperature at the soil surface is observed under conditions of maximum insolation shortly after noon (around one o'clock), and the minimum just before sunrise. When penetrating into the soil there is a reduction in the amplitude of oscillation and a gradual shift of the maximum and minimum to later hours. This is well illustrated by Figure 70, which shows the results of observations on two successive days at different depths in the soil, from data of the Koltushi station (near Leningrad). One can see that while the amplitude is about 30° at the soil surface, at a depth of 20 cm it amounts to only 2–3°. At the same time the maximum, observed near the soil surface at about 14 hrs, occurs at about midnight at a depth of 20 cm.

The decrease in the amplitude of daily oscillations with depth and delay in the arrival of extremum values may be judged from the data presented in Table 45.

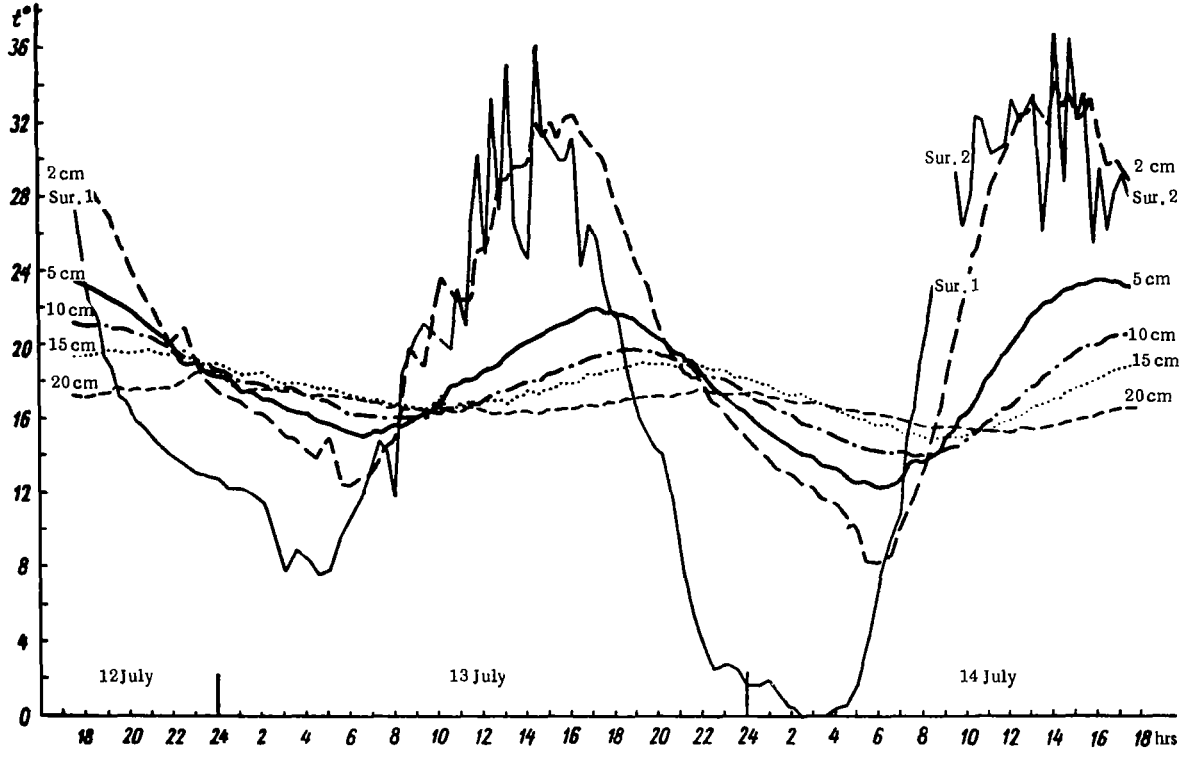


FIGURE 70. Distribution of temperature at soil surface (sur.) and at depths of 2, 5, 10, 15, and 20 cm. Koltushi, 12-14 July 1948

We note that the amplitude and depth of penetration of temperature waves depend largely on the thermal properties of the soil and are different for different soils. On the average the depth of penetration of the daily temperature wave varies between about 35 and 100 cm depending on the properties of the soil and geographic conditions. The delay in the occurrence of maximum and minimum values amounts to 2.0–3.0 hrs on the average for every 10 cm of depth.

TABLE 45

Daily oscillations of soil temperature at Pavlovsk (annual mean)

Depth (cm)	Amplitude (°)	Time of occurrence (hrs)	
		maximum	minimum
Soil surface	14.1	13.2	3.4
20	2.7	18.2	8.1
40	1.0	23.7	12.8
80	0.2	7.0	19.0

The annual march of the soil surface temperature in middle and high latitudes is characterized by a maximum in July or August and minimum in January or February. In tropical regions the pattern is more complex owing to the effect of clouds. The amplitude of the annual march on bare surfaces is fairly large and greater in low than in high latitudes. Observations show that on the yearly average the soil is always warmer than the air by several tenths of a degree. The amplitude of the annual wave decreases with depth but much more slowly than that of the daily wave. Relevant data are presented in Table 46 for Tbilisi.

TABLE 46

Amplitude of annual oscillations of soil temperature at Tbilisi

Depth (cm)	1	20	40	84	165	326	399	647
Amplitude (°)	32.5	29.1	26.0	20.7	13.9	3.8	4.0	1.5

The depth of penetration of the annual wave, which should theoretically be 19 times greater than that of the daily wave, is actually considerably larger than this. In low latitudes it amounts to about 5–10 m and in middle and high latitudes to 8–24 m (up to 30 m in certain cases). The delay in penetration of the annual temperature wave averages 20–30 days per meter of depth.

The very interesting question of the maximum soil surface temperatures observed over the globe can only be answered approximately. Single observations in the deserts of the hottest regions of the earth give values reaching 70° and somewhat higher. Thus a temperature of 71.5° was observed at Tucson (Arizona) and 69° at Agra (India). The amplitudes of daily waves, which are also large, reach values of the order of 56.5°; in tropical regions they usually amount to about 25° on clear days.

§ 5. Influence of plant cover and other factors on the thermal regime of the soil

The presence of a plant cover on the soil surface has an important influence on the thermal regime of the soil. First of all, the plant cover casts a shadow on the surface, thus reducing the warming of the soil in daytime by solar radiation. In the night hours the plant cover reduces the cooling of the soil surface by trapping the heat emitted by the soil.

In addition, in the presence of a plant cover there is an increase in the loss of incident heat by evaporation of moisture; also, heat is lost in the growth process. Therefore a soil from which the natural vegetation has been removed (bare soil) will be hotter in summer in the hours of daylight than a plant-covered soil. Results of the measurement of soil temperatures at Leningrad (Lesnoi) on plots with and without natural cover (vegetation in summer and snow in winter) are given in Table 47 (February and July).

TABLE 47

Temperature difference (in degrees) between bare soil and soil with natural cover. Leningrad (Lesnoi)

Months	Depth (cm)				
	0	20	40	80	160
February	-7.16	-6.36	-5.44	-3.91	-1.62
June	4.27	3.91	3.41	2.23	0.55

As one can see from the table, in June the bare soil is warmer at all depths than the grass-covered soil, whereas in February, when the soil is covered by snow, the opposite obtains.

Snow crops have a similar influence on soil temperature. For example, measurements among sowings in the Kalinin region revealed that at a depth of 10 cm in June the soil was 6° cooler under wheat, oats and clover than under a fallow: the soil under a rye crop was 3° cooler than the soil under a fallow.

Forests also affect the variation of the thermal regime of the soil. In summer the soil in a forest is cooler than in a field or clearing. The opposite relationship is observed in winter but the temperature gap is smaller than in summer. On the yearly average forest soil is cooler at all depths by 2-2.5° than field soil.

Cultivation of the siol (plowing and loosening) are of considerable importance for the thermal regime of bare soils. Cultivation reduces the thermal conductivity of the soil and slows down heat transfer. As a result both heating and cooling of the surface are considerably stronger in the case of plowed or loosened soil than in the case of unplowed and unloosened soil. The temperature difference between cultivated and uncultivated soil can sometimes reach 5° and more. Even the difference in mean temperatures over a summer can amount to nearly 2°.

Irrigation, which also has a considerable influence on heat transfer in the soil, usually causes a reduction in the surface temperature by several degrees.

Local topography is also very important in this respect, largely due to differences in the influx of heat on slopes with various aspects. Sometimes the difference in surface temperature between northern and southern slopes can reach several degrees.

§ 6. Snow cover and its significance for the thermal regime of the soil

The snow cover is an especially important factor in the thermal regime of the soil, a fact pointed out already by A. I. Voeikov.

The calorimetric conductivity of snow is very low and depends largely on its density. The low conductivity of snow leads to a pronounced attenuation of the heat transfer between soil and atmosphere. In this way the snow cover shields the soil from deep freezing and sharp temperature changes. The temperature of soil surfaces under snow is usually higher than the surface temperatures of either snow or bare soil. As a result the mean annual temperature of the uppermost layers of the soil is considerably higher than the mean temperature of the air (by 2.3° at Moscow, 3.1° at Perm, 6.3° at Sakhalin, etc.).

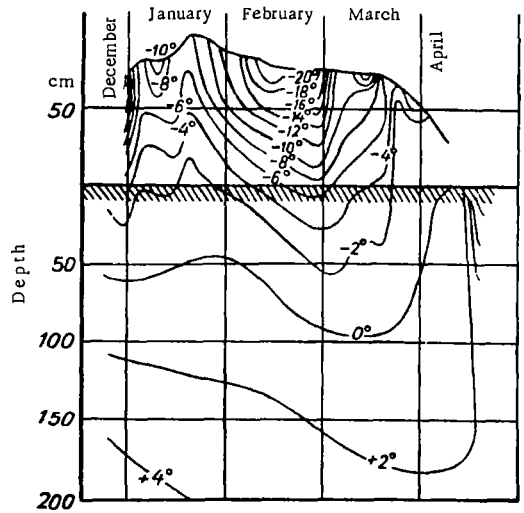


FIGURE 71. Isotherms in snow cover and soil. Borovoe experimental forest station, winter 1909–1910

Also important are the radiative properties of snow. The snow surface is characterized by very high values of the albedo for short-wave radiation and its emissivity is close to that of a black body. At the same time solar radiation penetrates deeply into the snow cover, in some cases even reaching to the soil surface (when the snow cover is thin). This explains, on the one hand, the fact that the snow cover is heated not only at the surface

but also in the interior, and, on the other hand, the strong cooling of the snow surface, which is always cooler than the adjoining layers of air.

The vertical distribution of temperature in snow and in snow-covered soil, and also the march of temperature at various depths in the snow cover and soil, can be analyzed with the aid of isotherms. An example is given in Figure 71. The uppermost line characterizes the variation of the thickness of the snow cover over the winter. One can see that the largest vertical temperature gradients occur in the snow. The largest temperature oscillations are also observed in the snow cover.

Investigations show that the daily march of temperature is felt only in the uppermost part of the snow cover, extending to a depth of about 25 cm. Even in this thin upper layer there is a very rapid falling-off with depth of the amplitude of the daily wave. All this convincingly demonstrates the protective role of the snow cover.

If in winter the snow cover has a warming effect on the soil, in the spring months it delays warming by screening the soil from solar radiation and expending large amounts of heat in thawing. As a result soil under snow is at a lower temperature in spring than bare soil.

§ 7. *Freezing of soil. Permafrost*

The heat stored by the soil in summer is returned to the atmosphere as winter approaches and the influx of solar radiation wanes. In winter the soil temperature drops below 0° and the soil freezes to a certain depth. This kind of soil freezing is termed seasonal, as it terminates in spring.

The process of formation of the seasonal freezing and thawing of the soil is very complex and depends on many conditions.

The depth of freezing, which is different for different areas, is determined not only by the length of the frost period and severity of the freezing temperatures but also by the character of the soil, its humidity, the extent of the vegetation cover, the thickness of the snow cover and many other conditions. Varying combinations of these conditions, which always act as a complex, are responsible for large contrasts in the development of this phenomenon between different areas.

The influence of the thickness of the snow cover on the depth of soil freezing can be judged from the data in Table 48, which were obtained from observations in the central part of the Baraba steppe (Western Siberia).

TABLE 48

Influence of thickness of snow cover on depth of soil freezing

Plot No.	I	II	III
Thickness of snow (cm)	180	80	20
Depth of freezing (cm)	10–12	70	90

In addition the depth of soil freezing depends on the plant cover, the presence of which reduces this depth. The influence of the character of a soil and its humidity can also be seen from the observational data; swamps,

for instance, freeze to a lesser depth (up to 50 cm) than dry areas, peaty soils freeze less than meadow soils (to 90 cm) and the latter still less than chernozem-meadow soils (130 to 185 cm).

Freezing is usually more extensive on bare spots than on areas covered by vegetation, particularly in forests. Freezing in a forest depends on the composition of the stands, their age, their density and so on. Thus observations in the Shipov forest (Voronezh region) demonstrated that, on the average over a number of years, forest soil freezes to a depth of 11 cm, as opposed to a freezing depth of 33 cm in fields.

The depth of soil freezing also depends on topography; in particular, it is greater and smaller, respectively, on rises and dips of the terrain than on level surfaces. The depth of freezing is affected by reclamation of swampy areas, felling of trees and plowing of land. Other conditions being equal such measures increase the depth of freezing. As a result the geographic distribution of freezing depths is highly complex. As a rule freezing is greater in the north than in the south, greater on bare surfaces (steppe regions) and greater in the presence of a limited snow cover. Within the confines of the European USSR and Kazakhstan the depth of freezing is particularly large (about 2.5 m) in Eastern Kazakhstan (50° N. lat) and in the Chernyi Irtysh and Lake Zaisan area. In most other districts (south and south-west of the country) the depth of freezing is usually less than 50 cm.

In certain fairly large northern areas soil freezing is observed at far greater depths than indicated above. Moreover, at a certain depth beyond the active layer the soil does not thaw even in the warm season and the soil temperature remains negative throughout the year. Such prolonged freezing (from two years to millenia) of the soil is called permafrost. In the USSR permafrost occupies vast expanses in Siberia, the Far East and polar regions, amounting to 47 % of the entire area of the USSR. In Eastern Siberia in many districts permafrost is recorded to 50° N. lat, and its southern boundary coincides approximately with the −2° mean annual isotherm. Vast areas are covered by permafrost in North America as well.

Permafrost occurs both continuously along the vertical, from the lower boundary of the active soil layer to a certain depth, and in the form of layers alternating with layers of thawed soil. The thickness of permafrost — the distance between its upper and lower boundaries — varies widely, from 1–2 m to several tens and even hundreds of meters. For example, the thickness of the permafrost layer is 274 m at Amderma, over 143 m at Yakutsk, and 1–2 m or less near the southern limit of the permafrost region (at Taishet station and near Lake Kizi). The top of the permafrost layer lies at a depth of between several tens of centimeters and several meters. The principal factors determining this depth are the influx of solar radiation, the snow cover and the calorimetric conductivity of the soil. The presence of a snow cover results in increased soil temperatures and therefore in a lowering of the upper boundary of the permafrost layer. High calorimetric conductivities also tend to lower the permafrost level. Low calorimetric conductivities have the opposite effect.

The level of occurrence of permafrost is also considerably affected by vegetation. For example, permafrost occurs rather at a higher level under shady deciduous trees and dense shrubs than in the steppe or under a sparse forest.

If the depth of thawing is sufficient (of the order of 1 m) the presence of permafrost in the summer period does not interfere with the cultivation of

many farm crops. However, permafrost raises considerable difficulties in construction work owing to the formation of so-called knobs and ice patches. Such features of soil structure occur when layers of thoroughly wet soil which have thawed in summer become trapped between two layers of ice under strong pressure during the autumn frost, with the result that the upper layer of ice is deformed and sometimes broken.

By modifying the character and state of the uppermost layer of soil one can influence the level of the permafrost layer artificially. Thus by removing the grass (or moss) cover, removing vegetation and making the soil compact one can lower the level of the permafrost layer; the opposite effect may be achieved by covering the soil with a loose, poorly conducting material.

The origin of permafrost is not entirely clear yet. A number of facts (e.g., the discovery inside permafrost ground of undecomposed mammoth herds) indicate that the layer was formed as early as the ice age, i.e., many tens of thousands of years ago. The severe climate of the territory occupied by permafrost tends to preserve the permafrost layer even today.

§ 8. Propagation of heat in water

Conditions of propagation of heat in bodies of water (oceans, seas, lakes, and so forth) are substantially different from those in the soil. This is due not so much to differences in the conditions of absorption and emission of radiation as to the fact that in bodies of water the principal factor in heat propagation is turbulent mixing— and not molecular heat conduction as in the soil.

Water absorbs radiant energy intensively in the infrared region. At the same time the emissivity of water ($\delta = 0.96$) is similar to that of various natural surfaces ($\delta = 0.90-0.95$). Hence the thermal emission of water consists of the radiation given off by a very thin surface layer, and there is no essential difference between a water surface and the soil as far as absorption and emission of long-wave radiation are concerned.

A different situation obtains in the absorption of solar radiation (short wave). As observations show, the latter penetrates to considerable depths in water, and rays with short wavelengths (violet and ultraviolet) penetrate particularly deeply. This deeply penetrating radiation produces a marked thermal effect when absorbed in a certain layer of water several centimeters thick.

However, the principal factor in the propagation of heat in water is the mobility of the water particles. The latter is responsible, firstly, for the mixing of cold and warm masses of water under the influence of purely mechanical forces (e.g., wind) and, secondly, for convective heat transfer. As a result the layer in which the daily and annual temperature waves are recorded is considerably larger in bodies of water than in the soil. For example, marked daily oscillations are observed at a depth of 15--20 m in the ocean; annual oscillations are even observed to a depth of 200–300 m. At the same time the amplitude of the temperature wave (daily and annual) must necessarily be limited. Observations show that the amplitude at the ocean surface is usually several tenths of a degree in the daily wave and a

few degrees in the annual wave. In shallow basins of limited size the amplitude of the daily wave reaches greater values (of the order of several degrees) while the amplitude of the annual wave is then of the same order as for land (about 15–20°).

The smallness of the amplitudes in oceans and seas is due also to the fact that the heat capacity of water is considerably larger than that of soil and therefore for the same amount of incident energy the temperature effect will be many times smaller in water than in the soil. Partly for the same reason but mainly due to turbulent mixing, large bodies of water are capable of accumulating 20–30 times as much heat as land. For example, for the Baltic Sea the annual influx and outflow of heat amounts to about 52,200 cal/cm^2 as opposed to 1800 cal/cm^2 for soil in the Leningrad area (cf. § 3).

The propagation of heat in deep bodies of water may be described theoretically by equation (2) of § 2, replacing the coefficient a by the coefficient of turbulent (thermometric) conductivity. This problem has been solved by a number of authors but comparison of theoretical results with observational data is not always possible (especially with reference to the annual march) and is complicated by the heat transfer due to sea currents.

We will not dwell further on the temperature regime of water basins (for details see textbooks of physical oceanography and hydrology). We mention only an important point for meteorology, namely that large bodies of water have a warming effect in the cold season and a moderating effect in the warm season, so that air temperatures in their proximity tend to be somewhat lower in summer and higher in winter.

Chapter 12

HEAT TRANSFER IN THE ATMOSPHERE

§ 1. Fluxes of heat in the atmosphere

In addition to the heat transfer due to radiation R (see Chapter 10), heat is transported in the atmosphere together with moving particles (small quantities) of air. This process is termed convective transfer in the general case; we will denote the convective flux of heat by P_c. If we denote the heat content of a unit mass of air (expressed in thermal units) by

$$i = c_p T, \tag{1}$$

where c_p is the specific heat at constant pressure, we can then write the following expression for the flux P_c

$$P_c = i\rho v_n = c_p T \rho v_n. \tag{2}$$

Here v_n is the component of the air velocity along the normal to the surface under consideration and ρ is the density of air.

Since the convective flux has both a magnitude and a direction, it is often more convenient to consider the vector flux

$$\vec{P}_c = c_p \rho T \vec{v}, \tag{3}$$

which we will understand to mean the amount of heat transported per unit time across a unit surface perpendicular to the air velocity vector $\vec{v}$.

Since the product $\rho \vec{v} = \vec{\mu}$ represents the momentum density,

$$\vec{P}_c = c_p T \vec{\mu}. \tag{3'}$$

Let us estimate the order of magnitude of convective fluxes of heat under atmospheric conditions. Taking $\rho = 10^{-3}$ g/cm^3, $T = 280$°K and $v_n = 5$ m/sec, we find from (2) that

$$P_c = 0.24 \cdot 10^{-3} \cdot 280 \cdot 5 \cdot 10^2 = 2.02 \cdot 10^3 \text{ cal/cm}^2 \cdot \text{min},$$

i.e., P_c is approximately 1000 times larger than the flux of direct solar radiation at the top of the atmosphere.

Expressions (2) and (3) for the convective flux of heat do not allow for the presence of inhomogeneities in the moving flux of air. However, individual air particles (small volumes) may have slightly different values of the temperature, density and velocity, which can be higher or lower than the average values for the entire moving mass of air. Strictly speaking, therefore, expression (2) should be written as

$$P_c = \frac{c_p}{s} \sum_{j=1}^{N} \rho_j T_j v_j s_j = \frac{c_p}{s} \sum_{j=1}^{N} T_j \mu_j s_j, \tag{4}$$

where s is the area of the surface crossed by the flux and s_j is the transverse cross section of the j-th particle characterized by the values ρ_j, T_j, v_j; summation should be carried out over all N particles crossing the surface s per unit time.

In the presence of such inhomogeneities in the air mass the surface s will be continually crossed by air particles having different values of T_j and μ_j. This gives us a basis for dividing the convective flux into two components, which we will denote by P_{ad} and P_e:

$$P_c = P_{ad} + P_e. \tag{5}$$

The first of these terms, P_{ad}, can be imagined as the amount of heat which would be transported per unit time across a unit surface if every individual particle of air crossing this surface had the same temperature T_{av} and the same momentum μ_{av}, equal to the average values of these elements over all the moving air under consideration. Obviously, for the flux of heat due to the orderly motions of such air particles having an average heat content i_{av} and moving at a certain average velocity v_{av}, one can write

$$P_{ad} = c_p \rho_{av} T_{av} v_{av} = c_p T_{av} \mu_{av} \tag{6}$$

We will call this flux advective.

With regard to the second term P_e, recalling what was said in Chapter 3 concerning turbulent motions in the atmosphere, one can consider it as the part of the flux which is created by the fluctuations of the quantity μ, or turbulent mixing. This flux of heat, which is due to the indicated irregular turbulent motions, is called the turbulent or eddy flux.

In Chapter 3 we derived a general expression for the vertical turbulent flux of any quantity q, which is reduced to the form

$$P_{e,z} = -\frac{\rho}{s} \frac{\partial \overline{q}(z,t)}{\partial z} \sum_1^N w_j s_j (z - z_j) + \frac{\rho}{s} \sum_1^N w_j s_j \delta q_j + \frac{\rho}{s} \sum_1^N w_j s_j q_j''.$$

In the case of heat transfer the transported quantity is the heat content, i.e., $q = c_p T$. Considering the transport in a vertical direction, we write the above expression in the form

$$P_e = -c_p \overline{\rho} k_z \frac{\partial \overline{T}}{\partial z} + c_p \overline{\rho} \left[\frac{1}{s} \sum_1^N w_j s_j \delta T_j + \frac{1}{s} \sum_1^N w_j s_j T_j'' \right], \tag{7}$$

where $k_z = \frac{1}{s} \sum_1^N w_j (z - z_j) s_j$ is the coefficient of turbulence in the vertical direction and all other symbols are as previously.

However, in the case of heat transfer it is not always possible to disregard the terms in brackets as was done before. Indeed, assuming that the vertical displacement of an air particle takes place adiabatically, we obtain for the change in its temperature over the path from the level z_j, where its vertical velocity is zero, to the level z

$$\delta T_j = -\gamma_a (z - z_j). \tag{8}$$

We can then write the first term in brackets as

$$\frac{1}{s} \sum_{j=1}^N w_j s_j \delta T_j = -\frac{1}{s} \sum_{j=1}^N w_j (z - z_j) s_j \gamma_a = -k_z \gamma_a, \tag{9}$$

which is different from zero.

If, exploiting the analogy between the processes of molecular and turbulent transport, we disregard the second term in brackets in expression (7), then, with (9), the latter takes the form

$$P_e = -c_p\bar{\rho}k_z\frac{\partial\bar{T}}{\partial z} - c_p\bar{\rho}k_z\gamma_a = -k_z\bar{\rho}c_p\left(\frac{\partial\bar{T}}{\partial z}+\gamma_a\right),$$

or, since $\frac{\partial T}{\partial z}+\gamma_a=\gamma_a-\gamma=\frac{T}{\theta}\frac{\partial\theta}{\partial z}$, where θ is the potential temperature,

$$P_e = -k_z\bar{\rho}\,c_p\left(\frac{T}{\theta}\frac{\partial\theta}{\partial z}\right). \qquad (10)$$

Indeed, from the expression for the potential temperature $\theta = T\left(\frac{\bar{p}_0}{p}\right)^{\frac{\varkappa-1}{\varkappa}}$, we have, taking the logarithmic derivative $\frac{1}{\theta}\frac{\partial\theta}{\partial z}=\frac{1}{T}\frac{\partial T}{\partial z}-\frac{\varkappa-1}{\varkappa}\frac{1}{p}\frac{\partial p}{\partial z}$. From here, making use of the hydrostatic equation and equation of state and performing simple transformations we obtain

$$\frac{T}{\theta}\frac{\partial\theta}{\partial z}=\frac{\partial T}{\partial z}+\frac{\varkappa-1}{\varkappa}\frac{g\rho}{p}\frac{TR}{R}=\frac{\partial T}{\partial z}+\gamma_a,$$

$$\text{since}\quad \frac{\varkappa-1}{\varkappa}\frac{g}{R}=\gamma_a,\quad \text{and}\quad \frac{\rho RT}{p}=1.$$

If the level which we are considering is close to the 1000 mb pressure surface, then $\theta \approx T$ and

$$P_e = -k_z c_p \rho\frac{\partial\theta}{\partial z}, \qquad (10')$$

i.e., the vertical turbulent flux of heat is directly proportional to the gradient of the potential temperature θ.

It therefore follows that the flux P_e is negative ($P_e<0$) for a dry-stable stratification of the atmosphere ($\frac{\partial\theta}{\partial z}>0$ or $\gamma<\gamma_a$), zero ($P_e=0$) for a dry-neutral stratification ($\frac{\partial\theta}{\partial z}=0$ or $\gamma=\gamma_a$) and positive ($P_e>0$) only for a dry-unstable stratification ($\frac{\partial\theta}{\partial z}<0$ or $\gamma>\gamma_a$).

However, the analogy between the processes of molecular and turbulent transport is not complete, and the discarded term in square brackets is not really zero. This is a result of the presence of Archimedes accelerations. Indeed, at the level z_j, at which the vertical velocity of the j-th particle was zero for the last time before crossing the level z, its acceleration is determined not only by dynamical but also by thermal factors. For particles that are warmer than the medium, i.e., for which $T_j''>0$, the Archimedes acceleration is positive, and vice versa; this means that particles reaching the velel z from below ($w_j>0$) were, on the average, warmer than the medium at the initial levels z_j and therefore $\Sigma T_j''>0$. Thus the sum $\frac{1}{s}\Sigma w_j T_j'' s_j$ is essentially positive, i.e., in the case where the vertical gradient of the mean temperature $\frac{\partial T}{\partial z}$ is equal to the adiabatic gradient $\left(\frac{\partial\theta}{\partial z}=0\right)$, the turbulent flux will not be zero, as obtained above. Air particles moving upwards will on the average have a higher potential

temperature and air particles moving downwards a lower potential temperature, and the total flux will be positive, i.e., directed upwards.

Let us now divide the quantity $\frac{1}{s}\sum_{j=1}^{N} w_j s_j T_j''$ by another positive quantity $\frac{1}{s}\sum_{j=1}^{N} w_j s_j (z - z_j) = k_z$. For the resulting quotient, which has the dimensionality of a temperature gradient, we introduce the symbol

$$\frac{\sum w_j s_j T_j''}{\sum w_j s_j (z - z_j)} = \gamma' = \gamma_a - \gamma_{eq}. \tag{11}$$

The quantity γ_{eq} will be called the equilibrium temperature gradient.

Introducing expressions (9) and (11) into equation (7), we find

$$P_{e,z} = -c_p \bar{\rho} k_z \left(\frac{\partial T}{\partial z} + \gamma_{eq} \right). \tag{12}$$

From this one can see that when the temperature gradient is equal to the equilibrium gradient the vertical turbulent flux of heat is zero. Since the numerator and denominator in the left-hand side of (11) are positive we have $\gamma_{eq} < \gamma_a$; then for a dry-neutral stratification of the atmosphere, where $\frac{\partial T}{\partial z} = -\gamma_a$, the turbulent flux P_e is positive and not zero.

Thus the physical meaning of γ_{eq} resides in the fact that for $\gamma = \gamma_{eq}$ the turbulent heat flux vanishes.

The concept of the equilibrium temperature gradient was introduced by M. I. Yudin and M. I. Budyko on the basis of the considerations outlined above. We note, however, that the question of the actual values of the equilibrium temperature gradient still requires further clarification both theoretically and experimentally.

According to calculations by Yudin and Budyko, the mean value of the equilibrium temperature gradient in the lowermost 100-meter layer can be taken to be 6°/km, with an error of ± 1°/km; that is, within the limits of accuracy of observations, it is identical with the mean value of the vertical temperature gradient in the free atmosphere. However, it does not follow from this that the upward and downward turbulent fluxes of heat also balance each other on the average. In fact, for vertical temperature gradients greater than the equilibrium gradient the stability of atmospheric movements is less and the intensity of turbulent mixing is bound to increase. On the other hand, under conditions of high thermal stability of the atmosphere, especially when temperature rises with altitude, the intensity of vertical turbulent mixing decreases appreciably (sometimes by a factor of tens and hundreds). Downward turbulent fluxes of heat are therefore small. Calculations show that nearly all over the earth the total turbulent fluxes of heat are directed upwards — from the surface to the atmosphere. Obviously, if $\frac{\partial T}{\partial z}$ is very large by comparison with γ_{eq} (exceeds it by a factor of tens and hundreds) then, disregarding γ_{eq}, expression (12) can be written as

$$P_{e,z} = -c_p \bar{\rho} k_z \frac{\partial T}{\partial z}. \tag{12'}$$

In particular this holds in the very thin layer of air immediately adjoining the earth's surface.

We note that owing to the daily march of the coefficient of turbulence, in which its values increase in daytime and are small at night, the turbulent flux of heat from the earth's surface to the atmosphere is greater in daytime and the flux in the opposite direction (from atmosphere to surface) is small at night; this is the so-called valve effect, responsible for the fact that on the average the flux of heat is directed from the surface to the atmosphere.

If one considers the transport of heat in the horizontal direction (along the x-axis) then instead of equation (7) one can write the analogous expression

$$P_{e,\,x} = -c_p \bar{\rho} k_x \frac{\partial T}{\partial x} + c_p \bar{\rho} \left[\frac{1}{N} \sum_{j=1}^{N} u_j \delta T_j + \frac{1}{N} \sum_{j=1}^{N} u_j T_j'' \right], \qquad (13)$$

where

$$k_x = \frac{1}{N} \sum_{j=1}^{N} (x - x_j) u_j$$

is the coefficient of horizontal transfer.

Now the second term in brackets in (13) is rigorously zero, since the temperature difference between the particle and the medium does not produce horizontal accelerations. As to the first term, the quantity $\left(\sum_{j=1}^{N} u_j \delta T_j \right)$ is not zero even in this case owing to the possible adiabatic temperature changes during compression and expansion of the particles in horizontal displacements. However, these temperature changes are generally very small. Taking this into account, we obtain

$$P_{e,\,x} = -\bar{\rho} c_p \frac{T}{\theta} k_x \frac{\partial \theta}{\partial x}. \qquad (14)$$

For $P_{e,\,y}$ we obtain, analogously,

$$P_{e,\,y} = -\bar{\rho} c_p \frac{T}{\theta} k_y \frac{\partial \theta}{\partial y}. \qquad (14')$$

The same value of the coefficient of horizontal transfer $k_x = k_y = k_0$ can be used in both formulas, since the directions x and y are in general equivalent.

In deriving the preceding formulas the general definition of convection given in the beginning of the section was used. However, in the atmosphere the horizontal component of the velocity of air masses is usually considerably (hundreds of times) greater than the vertical component; at the same time (and this is especially important) in the horizontal direction the inhomogeneity of the flow is much less than in the vertical direction. In view of this one can say that the convective flux $P_c = P_{ad} + P_e$ in the horizontal direction is chiefly determined by P_{ad}, the horizontal component of which is called the advective heat flux. In meteorology the horizontal transfer of heat is therefore termed advective, and the corresponding process is known as advection. By convection is understood the transport of air particles in the vertical direction over heated areas of the earth's surface.

§ 2. Influx of heat and its relationship with the flux

The changes in the thermal state of a certain volume of atmospheric air are determined not by the [vector] flux of heat but by the [scalar] influx. The influx of heat $\varepsilon = \rho c_p \frac{dT}{dt}$ is the amount of thermal energy received from the outside by a unit volume of air per unit time. We will call this influx of heat to a fixed particle of air the individual influx. Together with this quantity one sometimes considers the local heat influx $\varepsilon' = \rho c_p \frac{\partial T}{\partial t}$, understood to be the amount of heat received from the outside per unit time per unit volume at a definite point in space. The influx of heat is measured in cal/cm^3 sec or in cal/cm^3 min.

Heat influx may take place due to the following causes: 1) absorption of radiant energy in the volume under consideration; we will call this the radiative influx and denote it by ε_r; 2) exchange of particles of the volume in question with particles of the surrounding air. The latter takes place in convection, and a distinction can be made between the turbulent influx of heat ε_e and the advective influx of heat ε_{ad}.

The influx of heat ε is related to the flux P by the simple expression

$$\varepsilon = -\operatorname{div} P, \tag{15}$$

i.e., the influx of heat is equal to the divergence of the heat flux taken with the opposite sign.

This relationship makes it possible to determine the influx of heat if the flux is known. Proving (15) is complicated in the general case and becomes considerably simpler in the particular case where the local influx of heat ε' is equal to the individual influx ε. This happens when the velocity of the volume in question is zero, in which case the advective influx ε_{ad} is also zero.

§ 3. Radiative and turbulent influx of heat

Let us use relation (15) to derive the expression for the radiative influx of heat ε_r. We recall that the derivatives of the radiant fluxes along the horizontal are many times smaller than the derivatives of these fluxes in the vertical direction, and therefore the former can be disregarded. This is connected with the fact that the atmosphere is considerably more homogeneous in the horizontal direction than in the vertical. Then, denoting by G and U respectively the total downward and upward fluxes of long-wave radiation and by S the total (downward) flux of short-wave direct solar radiation we obtain, on the basis of (15),

$$\varepsilon_r = -\frac{\partial}{\partial z}(U - G - S). \tag{16}$$

Recalling the discussion of radiative transport in the atmosphere in Chapter 10 and taking absorption into account, one can write the following approximate expression:

$$\varepsilon_r = \rho \sum_\lambda k_\lambda \left[U_\lambda - G_\lambda - 2\delta E_\lambda\right] + \rho' \sum_\lambda k'_\lambda S_\lambda \sec z_\odot, \tag{17}$$

where ρ' and ρ are the density of the absorbing substance (water vapor, carbon dioxide, etc.), k_λ and k'_λ are the monochromatic absorption coefficients for the wavelength λ, E_λ is the energy of black-body emission for the given wavelength, $z_\odot$ is the zenith distance of the sun and δ is the absorptance.

The summation in expression (17) extends over all wavelengths of short-wave (solar) and long-wave radiation. Next, in view of the presence of many substances which absorb and emit radiant energy in the atmosphere, we ought to write an expression analogous to (17) for each of these, choosing suitable values of ρ and k_λ, and then take their sum, which would give us the total influx ε_r. From this one can see how complicated the expression for the total influx ε_r becomes in the general case.

The expression for the turbulent influx of heat $\varepsilon_e = -\operatorname{div} P_e$ is no less complicated.

Indeed, recalling (10), (14) and (14'), one can write the expression for the turbulent heat influx to a unit mass

$$\varepsilon_e = -\operatorname{div} P = \frac{1}{\rho}\left(\frac{\partial P_{e,x}}{\partial x} + \frac{\partial P_{e,y}}{\partial y} + \frac{\partial P_{e,z}}{\partial z}\right) =$$

$$= c_p \frac{T}{\theta}\left[\frac{\partial}{\partial x}\left(k_x \frac{\partial \theta}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_y \frac{\partial \theta}{\partial y}\right) + \frac{\partial}{\partial z}\left(k_z \frac{\partial \theta}{\partial z}\right)\right]. \qquad (18)$$

Using expression (12) instead of (10), we obtain

$$\varepsilon_e = c_p \frac{T}{\theta}\left[\frac{\partial}{\partial x}\left(k_x \frac{\partial \theta}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_y \frac{\partial \theta}{\partial y}\right)\right] +$$

$$+ c_p \frac{\partial}{\partial z}\left[k_z\left(\frac{\partial T}{\partial z} + \gamma_{eq}\right)\right], \qquad (18')$$

where all symbols are as previously.

§ 4. Internal sources of heat

In addition to the fluxes of heat considered above, other factors can cause changes in the internal energy of the air mass under consideration. They include, first and foremost, 1) adiabatic changes of pressure, and 2) phase transitions of the water contained in air. Let us examine their influence on temperature variations.

The following relation holds in adiabatic changes of state (for one gram of air):

$$c_v \frac{dT}{dt} + Ap \frac{dv}{dt} = \frac{dQ}{dt} = 0. \qquad (19)$$

For a unit volume we have

$$c_v\left(\frac{dT}{dt}\right)_a = -Ap\frac{dv}{dt} = A\frac{p}{\rho^2}\frac{d\rho}{dt} \qquad (20)$$

or, on the basis of equation (4'') of Chapter 6,

$$c_p\left(\frac{dT}{dt}\right)_a = \frac{A}{\rho}\frac{dp}{dt}. \qquad (20')$$

Here the subscript 'a' indicates that the temperature change is due to an adiabatic change of pressure.

Turning to the second factor, we confine ourselves to considering the condensation of the water vapor contained in air. If r grams of the water vapor contained in one gram of air change into water per unit time, the quantity of heat released in this process is Lr, where L is the latent heat of vaporization. If r' g of water transform into ice, an additional amount of heat, given by $L'r'$, where L' is the latent heat of fusion, is released. The heat released in these phase transitions will raise the temperature of a unit mass of air by $\left(\frac{\partial T}{\partial t}\right)_{ph}$ per unit time, where

$$c_v\left(\frac{dT}{dt}\right)_{ph} = Lr + L'r'. \tag{21}$$

Here the subscript 'ph' indicates that the cause of the temperature change is a process of phase transition of water.

The heat received by air as a result of the phase transitions of moisture may also be treated as essentially a form of heat influx; we denote this influx per unit volume by ε_{ph}. Then, on the basis of equation (21), we have

$$\varepsilon_{ph} = \rho c_v\left(\frac{dT}{dt}\right)_{ph} = \rho\,[Lr + L'r']. \tag{22}$$

In addition to the two internal sources mentioned above, one might indicate a series of secondary processes capable of leading to an influx of heat to the air. For example, heat always flows in as a result of molecular conduction. It is possible to have a certain influx of heat due to the transformation of kinetic into thermal energy by friction, and so forth. In certain cases one has to take into account the heat influx to the earth's surface due to temperature differences between precipitation and the ground. However, all these influxes in the atmosphere, save for very rare exceptions, are so small that they can be disregarded in comparison with the basic factors considered earlier.

§ 5. Equation of heat influx

A general expression (4") for the influx of heat per unit mass in the atmosphere was obtained in Chapter 6, § 1. For a unit volume it is written as follows:

$$\varepsilon = \rho\,\frac{dQ}{dt} = \rho c_p \frac{dT}{dt} - A\,\frac{dp}{dt}. \tag{23}$$

Taking into account all the major sources of heat considered above, we can write (23) as

$$\varepsilon = \rho c_p \frac{dT}{dt} - A\,\frac{dp}{dt} = \varepsilon_r + \varepsilon_c + \varepsilon_{ph} = \varepsilon_r + \varepsilon_e + \varepsilon_{ad} + \varepsilon_{ph}. \tag{24}$$

Since $\rho c_p \frac{dT}{dt} - A\,\frac{dp}{dt} = \rho c_p \frac{T}{\Theta}\frac{d\Theta}{dt}$, where Θ is the potential temperature, the above equation can also be written in the form

$$\varepsilon = \rho c_p \frac{T}{\Theta}\frac{d\Theta}{dt} = \varepsilon_r + \varepsilon_e + \varepsilon_{ad} + \varepsilon_{ph}. \tag{24'}$$

Secondary sources are not accounted for in this equation, which is therefore approximate.

In view of the fact that the term $-A\frac{dp}{dt}$ in (24), which accounts for the heat influx due to pressure changes, is very small compared with the rest (this is easily shown) it may be disregarded. We then obtain the approximate equation of heat influx

$$\varepsilon = c_p\rho\frac{dT}{dt} = \varepsilon_r + \varepsilon_e + \varepsilon_{ad} + \varepsilon_{ph}. \tag{25}$$

In the absence of advection, when the term $\varepsilon_{ad} = 0$, this equation assumes the still simpler form

$$c_p\rho\frac{dT}{dt} = \varepsilon_r + \varepsilon_e + \varepsilon_{ph}. \tag{26}$$

In the right-hand sides of the above equations the terms ε_r, ε_e, ε_{ad} and ε_{ph} are given by the relations obtained earlier. Thus even in these forms the approximate equation for the heat influx is highly complex. It cannot be solved in the general form but is entirely suitable for the solution of a wide range of particular problems in which simplifications can be made. We will proceed to demonstrate this below.

§ 6. Advection of heat and cold

One of the terms of the equation of heat influx is the influx due to advection, ε_{ad}. Let us consider it in somewhat greater detail. Since advection produces nonperiodic temperature changes in the atmosphere above the boundary layer, in considering small time intervals one usually disregards all other forms of heat influx. The equation of heat influx can then be written in the form (equation (4"), chapter 6)

$$\varepsilon = c_p\frac{dT}{dt} - \frac{ART}{p}\frac{dp}{dt} = 0. \tag{27}$$

In view of the fact that the total individual derivatives appearing in this expression are

$$\frac{dT}{dt} = \frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} + w\frac{\partial T}{\partial z},$$

$$\frac{dp}{dt} = \frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x} + v\frac{\partial p}{\partial y} + w\frac{\partial p}{\partial z},$$

where u, v and w are the components of the velocity along the coordinate axes, we rewrite (27) as follows

$$\frac{\partial T}{\partial t} = -\left(u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y}\right) - w\frac{\partial T}{\partial z} +$$

$$+\frac{ART}{pc_p}\left(\frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x} + v\frac{\partial p}{\partial y}\right) + \frac{ART}{pc_p}w\frac{\partial p}{\partial z}. \tag{28}$$

The third term in the right-hand side is seen to be considerably smaller (at least ten times) than each of the remaining terms, and can therefore be disregarded.

The fourth term may, in view of the hydrostatic equation $\frac{\partial p}{\partial z} = -g\rho$ and the equation of state $\rho RT = p$, be brought to the form

$$w \frac{ART}{c_p p} \frac{\partial p}{\partial z} = -w \frac{Ag}{c_p} = -w\gamma_a, \tag{29}$$

where γ_a is the dry adiabatic gradient.

Bearing in mind that $-\frac{\partial T}{\partial z} = \gamma$, we write (28) as follows:

$$\frac{\partial T}{\partial t} = -\left(u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y}\right) - w(\gamma_a - \gamma). \tag{30}$$

The first term in this equation gives the advective temperature change, the cause of which is the horizontal transport of air masses (advection). We will denote it by $\left(\frac{\partial T}{\partial t}\right)_{ad}$.

Consequently,

$$\left(\frac{\partial T}{\partial t}\right)_{ad} = -\left(u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y}\right). \tag{31}$$

If the transfer of heat takes place from a region with higher temperatures to a region with lower temperatures, one speaks of the advection of heat; when the transfer is in the opposite direction we have the advection of cold.

The second term in equation (30) gives the temperature change due to the vertical component of air motion

$$\left(\frac{\partial T}{\partial t}\right)_w = -w(\gamma_a - \gamma). \tag{32}$$

It is easily seen that the temperature change $\left(\frac{\partial T}{\partial t}\right)_w$ may be both positive and negative at any level depending on the sign of w and the stratification of the atmosphere as determined by the quantity γ. In the case of the ascent or descent of moist saturated air the expression for $\left(\frac{\partial T}{\partial t}\right)_w$ is more complex (cf. Chapter 6).

Chapter 13

TEMPERATURE OF THE LOWER ATMOSPHERIC LAYERS

§ 1. Mean distribution of temperature at the earth's surface

Our information concerning the temperature of the lower atmospheric layers is derived chiefly from measurements conducted at meteorological stations at a standard height of 2 m above the ground. For more extensive study of the structure of the temperature field measurements are carried out at other heights as well (above 2 m and below). The measurements reveal a complex distribution of temperature in the layer under consideration and indicate the presence of considerable periodic (daily or annual) and nonperiodic variations.

Let us first of all consider the general character of the temperature field over the earth. In studying this question it is usual to take values averaged over a certain interval of time (day, month, season, year). Charts of isotherms are constructed from these data (isotherms are smooth curves drawn through points having the same mean temperature).

Figures 72 and 73 show global charts of isotherms reduced to sea level (or rather to $z = 2$ m above sea level) for January and July. The isotherms do not lie along parallels of latitude in either hemisphere; this feature is especially pronounced in winter in the northern hemisphere, where the direction of the isotherms over, say, the European USSR is practically meridional. In the southern hemisphere, more uniform and largely occupied by oceans, the isotherms are smoother and run more nearly east-west. The deviations of the isotherms from the parallels are due to low land temperatures coupled with comparatively high sea temperatures in winter, and to the opposite relationship in summer. The highest temperatures are recorded near the equator, but the zone of highest temperatures (so-called thermal equator) is shifted away from the geographical equator toward the northern hemisphere— to 5–10° lat in January and 20° lat in July. Thus the position of the thermal equator varies from winter to summer. Its general northward displacement can be attributed to the presence in the northern hemisphere of great land masses subject to strong heating in summer.

The spacing of the isotherms varies from winter to summer in both hemispheres. In the northern hemisphere isotherms are more closely spaced in January than in July; in the southern hemisphere the situation is reversed.

Both in winter and in summer the lowest temperatures for the southern hemisphere are observed in Antarctica. In the northern hemisphere the

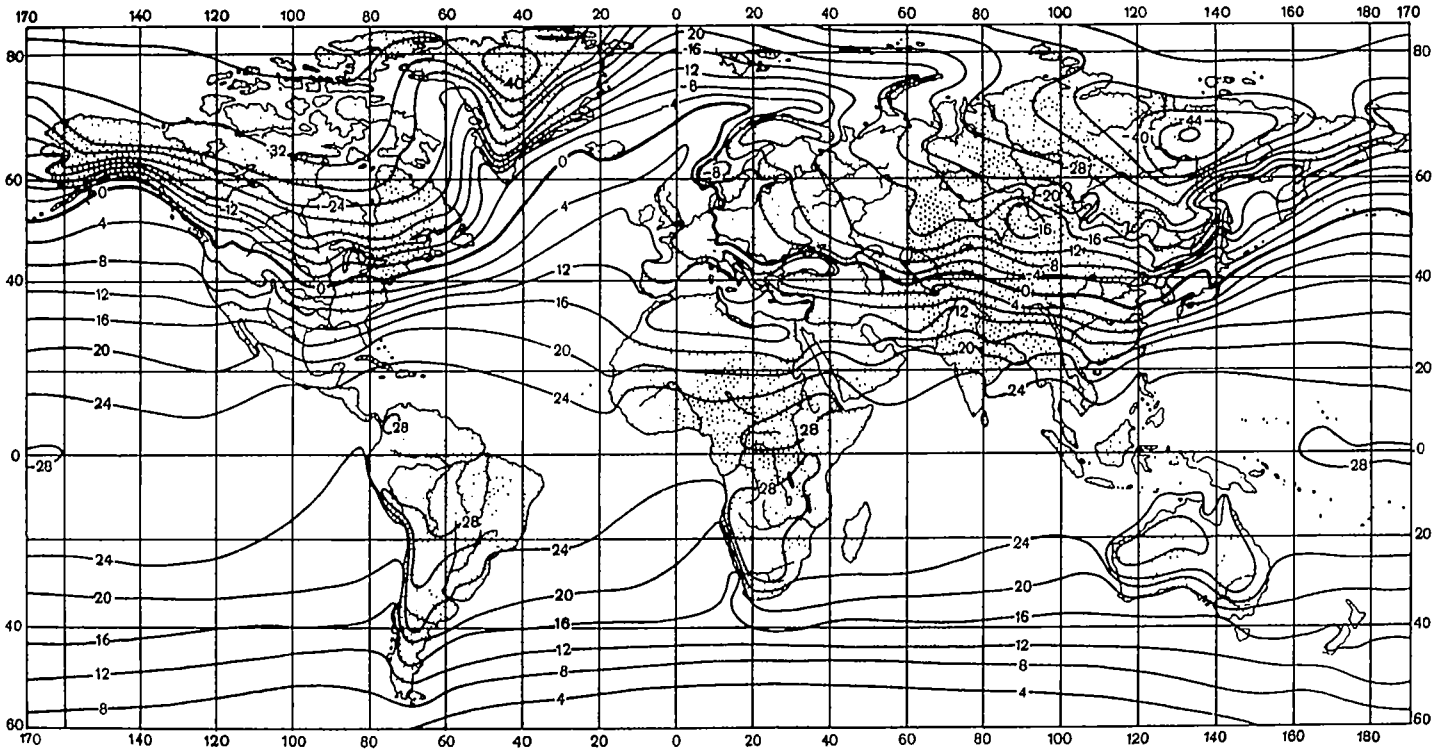


FIGURE 72. Isotherms (January)

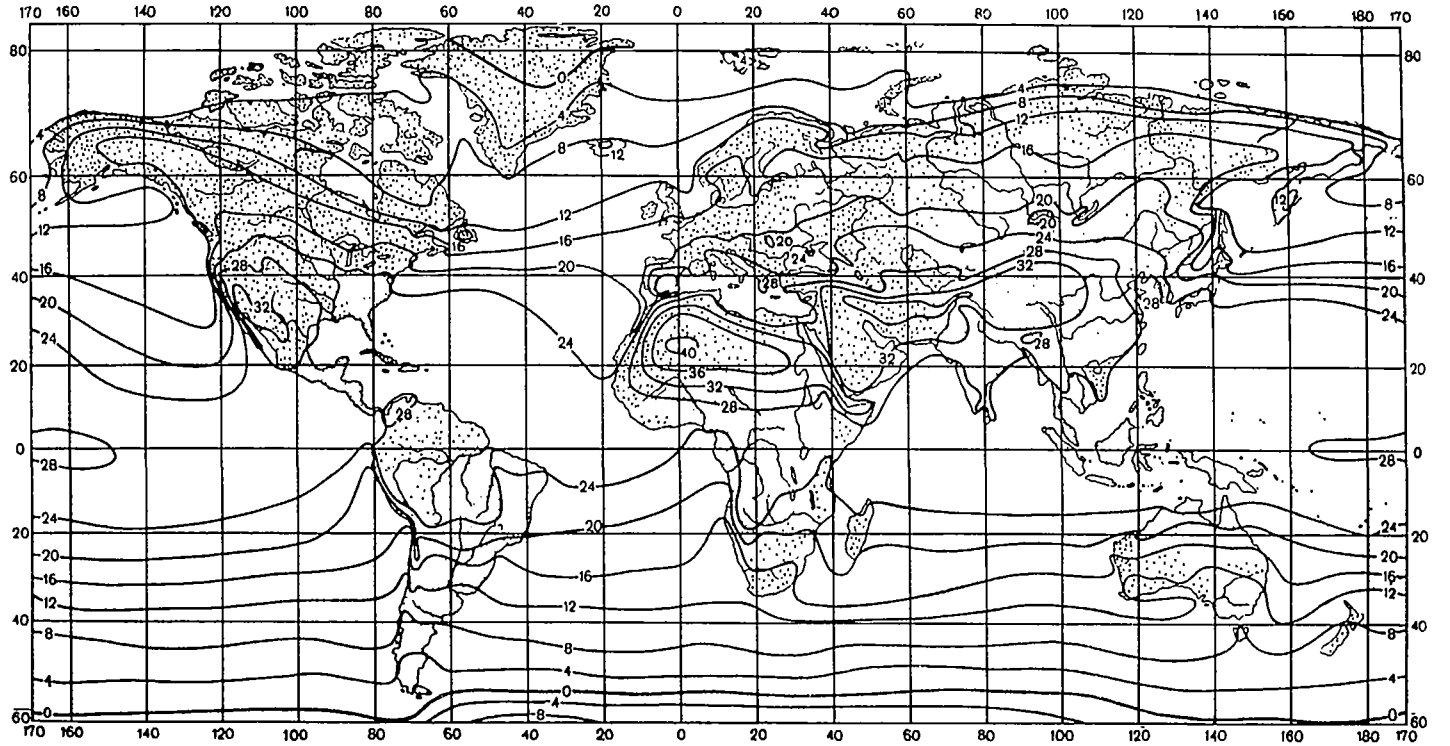


FIGURE 73. Isotherms (July)

coolest zone lies near the pole only in summer, and the polar regions are not the coldest in winter. The January charts of the northern hemisphere show two low temperature regions, known as the poles of cold, and both lie south of the pole. The principal one is found in USSR territory, near Verkhoyansk and Oimekon (Yakutia). Mean January temperatures are below −48° in this area, and below −40° in Yakutia as a whole. The second cold pole, with a mean January temperature of about −40°, lies over Greenland.

The lowest air temperatures recorded near the earth (at a height of 2 m from the surface) in the northern hemisphere have been below −60°; extremum values (absolute minimum) were recorded in Yakutia (−68° at Verkhoyansk, −70, −71° at Oimekon). Even lower temperatures have been recorded in the Antarctic, where values below −70° have been reported repeatedly; temperatures as low as −88° were reported at Vostok and Sovetskaya Stations in 1959.

The highest global temperatures are observed in the deserts of the hot tropical belt (15–40° N. lat). Mean July air temperatures in the Sahara, Arabia, Iran and Southern California are above 30°, with maximum temperatures exceeding 50°. The absolute maximum temperature can apparently be taken as 58°. This value was reported in Asia (Tripoli)* and in southern Iran. A temperature close to this, namely 56°, was recorded also in Death Valley, California, and in French Somaliland. Temperatures of about 50° have been observed in the Central Asian deserts of the USSR.

Thus the overall range of variation of air temperatures near the earth's surface (at a height of 2 m) amounts to about 140° (roughly from −80 to +60°)

A more detailed study of the isotherm charts will show that a moderating influence— warming effect in winter and cooling effect in summer— is exercised not only by vast surfaces of water such as the oceans but also by much smaller surfaces. An example is Lake Baikal, over which temperatures are higher in winter and lower in summer than in the surrounding areas.

The influence of warm and cold ocean currents is also apparent from the isotherm charts shown above.

A view of the zonal distribution of temperature may be obtained from the available climatological data by averaging the temperatures over parallels of latitude. Table 49 gives the mean sea-level temperature and pressure.

On the average, as the data in the table show, the thermal equator lies in the northern hemisphere, at 10° lat. The temperature gap between the thermal equator and the poles is different in different hemispheres and different seasons. In the northern hemisphere it amounts to about 29° in summer (July) and 68° in winter (January), and in the southern hemisphere to about 40° in summer (January) and 76° in winter (July). Thus the gap is particularly large in winter, nearly twice as much as in summer. Also noteworthy is the increase in the amplitude of annual temperature oscillations from the equator (where it is of the order of 1°) to the poles (35–40°). Comparison of the annual amplitudes at equal latitudes shows that they tend to be greater in the northern than in the southern hemisphere. This is due to the fact that land masses occupy a greater area in the former. In fact, land masses account for 39% of the entire area in the northern hemisphere, as against 19% in the southern hemisphere. As a result the

* [The Tripoli in North Africa is obviously intended here.]

northern hemisphere is warmer than the southern hemisphere at all latitudes if one considers the yearly average; the mean annual temperature of the northern hemisphere amounts to 15.2°, as opposed to 13.3° for the southern hemisphere and 14.2° for the earth as a whole.

TABLE 49

Zonal mean-monthly and mean-annual values of air temperature and pressure at sea level

Latitude	Air temperature (°)			Pressure (mb)			Difference, Jan-July		Difference, northern-southern hemisphere (per annum)	
	Jan	July	Annum	Jan	July	Annum	temperature (°)	pressure (mb)	temperature	pressure (mb)
90°N	—41.0	—1.0	—22.7	1013.5	1010.5	1015.0	—40.0	+8.0	+10.4	+23.9
80	—30.3	+2.0	—17.2	1012.5	1011.1	1014.2	—32.3	+1.4	+9.8	+24.5
70	—24.8	7.3	—10.7	1012.4	1010.3	1012.2	—32.1	+2.1	+2.9	+23.1
60	—15.6	14.1	—1.1	1014.3	1010.2	1011.5	—29.7	+4.1	+2.3	+22.7
50	—6.8	18.1	+5.8	1016.3	1011.9	1014.2	—24.9	+4.4	0.0	+10.0
40	5.5	24.0	14.1	1018.4	1013.3	1015.9	—18.5	+5.1	+2.2	+2.0
30	14.8	27.3	20.4	1019.3	1012.4	1015.5	—12.5	+6.9	+2.0	—2.4
20	22.1	28.0	25.3	1015.8	1010.5	1012.2	—5.9	+5.3	+2.4	—3.3
10	25.1	26.9	26.7	1011.9	1010.2	1010.5	—0.8	+1.7	+1.4	—1.7
0	26.7	25.6	26.2	1010.3	1011.9	1010.5	+1.1	—1.6		
10°S	26.5	22.9	25.3	1010.2	1014.5	1012.2	+1.6	—4.3		
20	25.8	20.0	22.9	1011.6	1017.9	1015.5	+5.8	—6.3		
30	22.9	14.7	18.4	1014.7	1020.3	1017.9	+8.2	—5.6		
40	16.1	9.0	11.9	1014.9	1014.5	1013.9	+7.1	+0.4		
50	8.8	3.4	5.8	1003.5	1003.9	1004.2	+5.4	—0.4		
60	2.1	—9.1	—3.4	989.6	988.3	988.2	+11.2	+1.3		
70	—3.4	—23.0	—13.6	990.9	989.2	989.1	+19.6	+1.7		
80	—10.8	—39.5	—27.0	992.0	990.2	990.7	+28.7	+1.8		
90	—13.5	—48 0	—33.1	992.6	991.1	991.1	+34.5	+1.5		

The departures of the real temperature distribution from the zonal one and in particular the role of continents and oceans in the temperature regime of the atmosphere can be estimated with the help of charts of temperature anomalies. These are obtained by calculating the differences between the zonal temperatures and the mean temperatures at any given point of the parallel, plotting these differences on a chart and joining points with the same values of the difference by smooth curves.

Such charts show that in January, for instance, air temperatures in the western USSR, on the east coast of the warm Atlantic, are more than 10° higher than the zonal temperature; in the Verkhoyansk area, on the other hand, the actual temperature is 26° below the zonal temperature. The opposite relationship is observed in July, though not so strongly expressed.

§ 2. Annual march of air temperature

The annual oscillations of air temperature at any point are determined primarily by the annual variations of the influx of heat from the sun. However, they depend also on the nature of the surface and other factors indicated earlier.

In the northern hemisphere the highest mean monthly air temperatures are recorded in July-August and the lowest in January-February. The influence of land and sea is felt in the fact that the amplitude of annual oscillations is considerably larger over continents than over oceans and shores; also, extremum values are observed somehwat earlier (nearer the solstices) over continents.

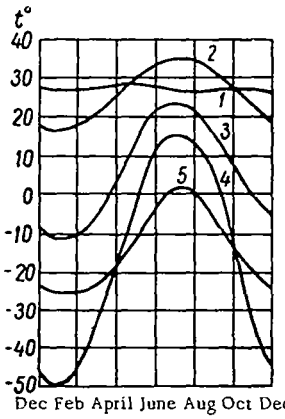


FIGURE 74. Annual march of temperature at different latitudes

1–Batavia (φ = 6.2° S); 2–Aswan (φ = 24° N); 3–Saratov (φ = 51° N); 4–Verkhoyansk (φ = 67° N); 5–Traurenberg (φ = 80° N).

Four types of annual march of air temperature may be distinguished according to latitude:

1. Equatorial type. In the equatorial zone, where the influx of solar radiation changes only slightly over the year, the amplitude of the annual temperature march is small: it amounts to about 1° over oceans and 5–10° over continents. Further, two maxima are recorded during the year— in the periods of the vernal and autumnal equinox— and also two minima— during the winter and summer solstice (Figure 74, curve 1).

2. Tropical type. As one moves away from the equatorial region there is an increase in the difference in solar heat influx between summer and winter, and therefore also an increase in the amplitude of temperature oscillations. Thus in the tropics the annual amplitude amounts to about 5–10° over oceans and about 15–20° over continents; there is one maximum after the summer solstice and one minimum after the winter solstice (Figure 74, curve 2).

3. Temperate zone type. Middle latitudes also have a simple temperature march with a maximum after the summer solstice (July) and minimum in winter (January). The annual amplitude here is high— 10–15° over the open sea and ocean, up to 40–50°, and even 60°, over continents, especially in regions distant from the sea (Figure 74, curves 3 and 4). The occurrence of the maximum and minimum over the sea is delayed: the maximum is shifted from July to August and the minimum from January to February.

4. Polar type. Polar regions are characterized by a long cold winter and comparatively warm but brief summer. The total annual amplitude is also large. The maximum occurs in August and the minimum at the end of the polar winter (March; see Figure 74, curve 5).

Monsoon regions have a rather special annual temperature march, with low temperatures in winter and summer and higher temperatures in late spring and autumn.

Further details and peculiarities of the annual march which depend on local conditions (elevation, relief, etc.) will not be considered here.

§ 3. Daily march of air temperature

We will first consider the principal characteristics of the daily march of temperature as revealed by observations at the standard height of 2 m.

As in the case of the annual march, the controlling factor in the daily march is the influx of heat from the sun to the surface. Heat is transmitted from the surface to the air layers, causing air temperatures to increase after sunrise (however, this increase lags behind the surface temperature and the lag increases with altitude). The highest temperatures are recorded 2–3 hours after midday; after this temperatures begin to drop, comparatively rapidly at first (until sunset) and slowly later (during the night). The minimum is reached shortly before sunrise. This is the general pattern of the daily march of air temperature. Details vary somewhat with latitude and season.

Due to the fact that the height of the sun at noon increases away from the poles, daily amplitudes are largest in subtropical latitudes and decrease polewards. At the same time the sun's height at noon changes only slightly during the year in equatorial regions but varies a great deal in middle and especially in high latitudes. In high latitudes, therefore, the daily amplitude should be greater in summer than in winter. However, amplitudes are determined not only by the value of the daytime temperature maximum but by the nighttime minimum as well; while the maximum depends on the solar elevation at noon, the minimum depends on the degree of nighttime cooling, which increases with the length of the night. In short summer nights, during which the earth's surface and therefore also the air are not appreciably cooled, the temperature minimum is not severe. As a result the largest daily amplitudes do not coincide with the summer solstice but are observed near the equinox, especially in high latitudes. At the same time daily amplitudes are subject to variation over the year. Thus in polar regions daily oscillations are practically absent in winter and are small in summer. The largest amplitudes recorded here occur in periods near the equinoxes. In middle latitudes the smallest amplitudes are observed in winter and the largest in early summer; amplitudes are slightly larger in spring than in summer. In tropical regions the daily oscillations are nearly constant in magnitude over the year and are greater than in middle latitudes.

Daily oscillations of air temperature are influenced by the properties of the underlying surface, on which depend, as we saw earlier, the degree of daytime heating and nighttime cooling of the surface. This influence is felt in the fact that daily amplitudes are limited over seas and oceans (to 2–3°) but are considerably larger over continents (to 20–22°). A similar influence is exerted by smaller bodies of water (lakes) and wet surfaces (swamps, spots with abundant vegetation), which reduce the amplitude of daily oscillations; bare dry areas (especially in deserts) tend to increase it.

In winter the presence of a snow cover increases the daily air temperature amplitudes due to the influence of reflection and emission. Obviously enough, daily amplitudes also depend on cloudiness: the daily march on cloudy days is very smooth compared with that on clear days. When there is variable cloud, which reduces both the daytime warming and the emission of the earth's surface, the daily march may depart substantially from the pattern described.

In addition to these factors the daily march is influenced by many other physico-geographical conditions. Thus, A. I. Voeikov pointed out the strong influence of relief and elevation of the terrain. In the case of concave relief (valleys, troughs) there is greater daytime heating and nighttime cooling, and hence the amplitude is greater. In the case of convex relief (hills, mountains) the earth's surface has less influence on air temperature

and the amplitude is smaller. The daily march is also very sensitive to the horizontal transport of air masses, in particular the local circulation (breezes, mountain and valley winds and so on).

No less important are the differences in temperature regime due to vegetation. Plant covers display a highly complex distribution of temperature, and one finds a reduction in both the temperature and amplitude of oscillations in the air above them.

What we have said applies to the daily march of temperature at the height of standard meteorological measurements (2 m). Relationships change as one moves to other levels, especially to layers situated below 2 m. Daily oscillations are particularly large very close to the earth's surface. As one moves away from the surface the daily amplitude decreases and there is an increasing delay in the occurrence of the maximum and minimum. Observations show that daily oscillations are already barely perceptible at a height of 0.5 km in winter, and extend to a height of 1.5–2 km in summer. Higher up in the free atmosphere, in the middle and upper troposphere (above 2 km), one also finds small daily oscillations; these will be considered later.

The question of the daily march of temperature directly above the surface, where plants grow, is of particularly great interest.

Empirical analysis of the daily march is extremely difficult owing to the presence of numerous factors which act simultaneously and interact with each other in complex ways; therefore the observed regularities can be explained only in a qualitative manner. To establish the principal quantitative characteristics of the daily march in the surface layer of air one must attempt a theoretical solution of the problem. This was first done by A. A. Dorodnitsin and later by M. E. Shvets. In solving the problem they started with the equation of heat influx. Since the principal factor in the variation of the temperature regime is the turbulent flux of heat and the role of other fluxes is generally insignificant in the surface layer, in the first approximation one may disregard all fluxes other than the turbulent one. The equation of heat influx can then be written as follows:

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z}\right). \qquad (1)$$

In this form the equation of heat influx is called the heat equation of the atmosphere. This equation was employed by Dorodnitsin in his solution of the problem under consideration.

In addition to the vertical turbulent flux Shvets also took into account the radiative transport of heat in the atmosphere. Recalling equations (10'), (17) and (24) of Chapter 12, we obtain the equation of heat influx

$$c_p\rho \frac{\partial T}{\partial t} = c_p\rho \frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z}\right) + \rho k\,[U + O - 2\delta E] + k'\rho' s \sec z_\odot. \qquad (2)$$

Shvets did not include the frequency dependence (long- and shortwave) of the absorption of radiation in the atmosphere in his calculations. The solution of the given equation is discussed in textbooks of dynamic meteorology and cannot be considered here. Analysis of this equation shows that the vertical variation of air temperature is far more complex than the simple power law which holds for the variation of the soil temperature with depth. Furthermore, theoretical formulas indicate that turbulent transfer plays an important role: turbulence slows down the vertical fall-off of the

amplitude of the daily march and reduces the time lag in the occurrence of the maximum and minimum. In the case of very well-developed turbulent transfer daily temperature oscillations may extend to considerable altitudes.

We note that the theoretical formulas obtained were employed to compile tables and graphs for the computation of the daily march of temperature as a function of latitude, time of year, type of soil (i. e., albedo, thermal properties and so on) and atmospheric transmission.

Comparison of calculated and observed data reveals good agreement. It should be mentioned, however, that the above theory does not consider, and does not allow for, the very strong and complex influence of cloud.

§ 4. Vertical variation of temperature in the surface layer of air

The daily march of air temperature in the surface layer is closely linked to its vertical variation at different hours of the day. Earlier we mentioned that a characteristic feature of the surface layer is the presence of very

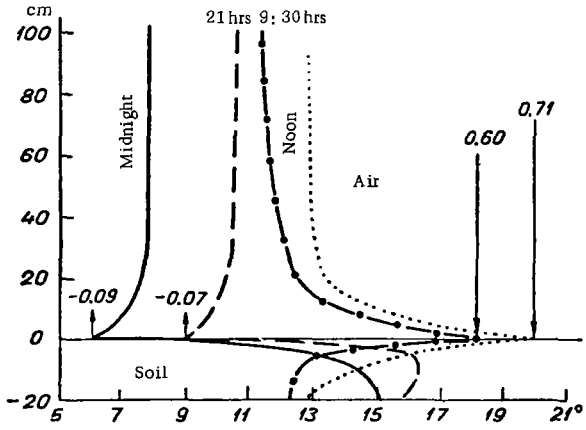


FIGURE 75. Vertical distribution of air temperature in surface layer (Koltushi)

steep (large absolute magnitude) vertical temperature gradients. The latter arise when large changes in the surface temperature are accompanied by feeble turbulent transport. During periods when the surface is being heated gradients many times greater than the adiabatic are formed in the atmospheric layers adjoining the surface; equally steep gradients in the opposite direction (inversions) appear when the surface and hence the adjoining layers of air are being cooled.

The temperature distribution in the surface layer for heating and cooling of the surface is illustrated graphically by Figure 75, which gives the results of observations at Koltushi (Leningrad area) on a clear day in May 1954. The numbers on the arrows show the values of the radiation balance

in cal/cm^2·min. Figure 76 shows a temperature profile observed at Leafield (southern England) over a meadow in spring (March); the profile ranges over a layer between 1.2 and 87.7 m. One can see from the figures how the vertical profile changes during the day. At night there is an inversion while in daytime, as usual, the temperature decreases with altitude. In the morning (between 8 and 9 hrs), soon after dawn, a state approaching isothermy is observed throughout this layer. A similar period of very gentle gradients occurs shortly before sunset.

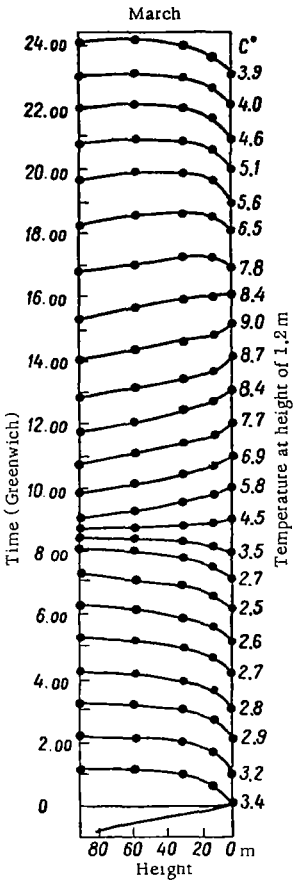


FIGURE 76. Air temperature profiles in the layer between 1.2 and 87.7 m at Leafield (southern England) in spring

Naturally, the time of disruption and formation of surface inversions coincides almost exactly with the time at which the turbulent flux of heat changes sign. In the lower 1.5-meter layer this happens at solar elevations of 10–15° on open ground, and even at solar elevations of 20–25° in the presence of a snow cover.

Comparing the temperature profile with the dry adiabatic (Figure 76) one can see that, to a certain height, the temperature gradient is considerably larger in absolute magnitude than the adiabatic. The steepness of gradients at different elevations can be judged from the data in Table 50, which gives the mean noon gradients in various layers based on observations at Porton (southern England).

The amount by which the observed gradients exceed the adiabatic may be seen from the numbers cited in the table.

The values cited for the gradients in the table are not maximal. Considerably steeper gradients sometimes appear over short time intervals; these may reach up to 1800 γ_a in the layer between 2.5 and 30 cm (again after Best's data for Porton).

The above statements apply to cloudless days. In the presence of cloud both heating and cooling of the surface become feebler. This results in milder vertical temperature gradients and a smoother daily march at all heights.

For a theoretical analysis of the vertical temperature profile in the surface layer we turn to the equation of heat influx. Recalling what was said on p. 237, we shall write this equation in its particular form (1), namely the heat equation of the atmosphere

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z}\right). \qquad (3)$$

We note that at the extrema (maximum and minimum) of the curve of the daily temperature march the derivative $\frac{\partial T}{\partial t}$ becomes zero; however, the

derivative $\frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z}\right)$ is also zero at these points. Since the vertical turbulent flux $P = -c_p \rho k_z \frac{\partial T}{\partial z}$, from what was said we obtain that in the surface layer

$$\frac{\partial P_e}{\partial z} = -c_p \rho \frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z}\right) = 0 \qquad (4)$$

or

$$P_e = -c_p \rho k_z \frac{\partial T}{\partial z} = \text{const.} \qquad (5)$$

This means that in the surface layer the turbulent flux of heat in a vertical column of air does not vary with altitude.

TABLE 50

Mean noon-temperature gradients over grass at Porton
(after Best)

Layer (cm)	$\frac{T}{T_a}$	
	January	July
25-30	100	625
30-120	11	78
120-710	2	14
710-1710	1	5

However, as was stated on p. 9 the coefficient of turbulence k_z varies linearly with height in the surface layer. Thus we can write

$$k_z = k_1 \frac{z + z_0}{z_1}, \qquad (6)$$

where k_1 is the value of the coefficient of turbulence at the height $z_1 = 1$ m and z_0 is the roughness length

Then equation (5) becomes

$$P_e = -c_p \rho k_1 (z + z_0) \frac{\partial T}{\partial z}. \qquad (7)$$

In view of the fact that $z_0 \ll z$, we write the above in the form

$$P_e dz = -c_p \rho k_1 z dT. \qquad (7')$$

Assuming that the temperature at the lower boundary of the layer z_0 is given by T_0 and denoting the value of P_e at this boundary by $(P_e)_0$, we integrate the above expression from z_0 to the height z at which the temperature is T

$$T - T_0 = -\frac{(P_e)_0}{c_p \rho k_1} \ln \frac{z}{z_0}. \qquad (8)$$

For any arbitrary level z_1 $(z_0 < z_1 < z)$ at which the temperature is given by T_1 we can write, from the above,

$$T_1 - T_0 = -\frac{(P_e)_0}{c_p \rho k_1} \ln \frac{z_1}{z_0},$$

from which

$$-\frac{(P_e)_0}{c_p \rho k_1} = \frac{T_1 - T_0}{\ln \frac{z_1}{z_0}}. \tag{9}$$

Introducing the above expression into equation (8) we find that

$$T = T_0 + (T_1 - T_0) \frac{\ln \frac{z}{z_0}}{\ln \frac{z_1}{z_0}}, \tag{10}$$

which gives us the law of variation of temperature with height in the surface layer. This formula is known as the logarithmic law of distribution of temperature with height in the surface layer. From it one can see that near the earth's surface (small z_1) the temperature changes very rapidly with height; further, if $T_1 > T_0$ then the temperature T increases with height, and vice versa. All the empirical regularities described earlier are in good agreement with this law.

§ 5. Frost

In some periods of the year (usually spring and autumn) the minimum temperature drops below 0° at certain hours of the day. A drop in the minimum to a value below 0° associated with a positive mean daily temperature is called frost. The temperature of the soil surface or vegetation during nighttime frosts may be considerably lower (2–3° or more) than the temperature of the air at a height of 2 m. Consequently, temperatures below 0° are sometimes recorded at plant surfaces even in the case of low positive air temperatures.

The study of frost is of great significance for many branches of the economy and especially for agriculture. In the latter context it should be recalled that for many cultivated plants frosts are destructive only when the temperature drops below definite critical values (up to several degrees), which are different for different plant species and growth stages. Frosts which are capable of ruining agricultural crops are known as killing frosts. The study of killing frosts is a special task of agricultural meteorology.

Frosts are grouped as follows according to origin:

a) radiation frosts, due to severe cooling of the soil surface and plants as a result of loss of heat by emission;

b) advection frosts, due to the influx into a given area of cold air masses from nearby regions (advection of cold), associated with a generally low temperature level.

Usually, however, both factors operate together, and the two types of frost are rarely observed in pure form.

All factors which assist radiative cooling of the active surface tend to favor frost. Such conditions are, as we saw, a cloudless (or nearly cloudless) sky, low humidity of the air and limited moistness of the soil surface, feeble turbulent transfer in the surface layer of air (this is related to low wind speeds, which drop at night to full calm), and unstable vertical distribution of temperature in the evening.

The probability of frost is greater in the presence of a large daily march of temperature, and also in areas characterized by a low calorimetric conductivity of the soil.

Frosts are experienced practically everywhere. Their duration is closely related to the daily amplitude of temperature and decreases as the latter increases; frosts may last several hours. For districts in the central and southern parts of the European USSR the duration averages about 4–5 hrs.

For prompt implementation of protective measures against the destructive effects of a frost it is necessary to predict its arrival ahead of time. Empirical rules of various kinds were devised for this purpose long ago; some are still in use. However, since such empirical laws usually relate frost only to some of the many factors that affect it (examples are the laws of Brounov, Mikhalevskii, Mikhel'son, Ångström, et al.) their accuracy is limited. Moreover, they are applicable only to the conditions under which they were obtained, and provide only crude estimates.

Significant improvements were made in the method of frost forecasting only after the question had been tackled on the basis of a physical analysis of the causes of night cooling, and, especially, after a theoretical solution had been obtained for the problem of the daily temperature march.

One of the earliest attempts to solve the problem is Brunt's formula, derived on the basis of the equation of heat conduction for the soil. This formula was modified to some extent in a number of subsequent works. However, the results obtained are rather unsuitable for practical calculations.

A more useful solution has been obtained by Soviet scientists and practical methods of frost forecasting have now been devised. A method proposed by A. F. Chudnovskii and I. G. Lyutershtein is based on a theoretical scheme of the daily temperature march derived from the solution given by M. E. Shvets for the daily march problem.

The next step forward was taken by M. E. Berlyand when he treated the nighttime temperature drop not as a part of the periodic daily march but rather as an aperiodic process for certain specific initial conditions He suggested a solution of the problem for specified effective emission depending on the initial distribution of temperature in soil and air, and also on the intensity of turbulent mixing. The complexity of the mathematical aspects prevents us from considering this theory in any detail.

By simplifying the complicated formulas obtained it has been possible to give simple computational formulas; these can be used for practical purposes to compute the air temperature at any height (including the soil surface) at any instant during the night.

With the aid of nomograms and computational tables the probability and character of a frost can be estimated with comparative ease and speed from data of ordinary meteorological observations. Verification of this method against a large volume of material shows very satisfactory agreement with observed data.

It should be stressed, however, that these methods are applicable only in the absence of advection, a factor which is very difficult to take into account.

Various methods dating back to a remote past are employed to protect farm plants from damage by killing frosts. The principal ones are:

a) open heating of plantations by means of heaters of various kinds and b) smudging with smoke. Individual and collective covering of plants is less often used (only for particularly valuable crops).

These methods are based on the principle of reducing nighttime cooling by influencing the factors which control the thermal regime of the surface layer.

Occasionally (especially in the republics of Central Asia) intensive wetting of the soil is used to combat frost; this measure ensures an increased flux of heat from the depths to the surface of the soil and partly helps to reduce effective emission.

§ 6. Thermal convection in the atmosphere

The vertical distribution of temperature in the surface layer of air is of considerable significance for the development of so-called thermal convection.

In theoretical and laboratory investigations of the question of heat transfer in the motion of liquids or gases (convection), a distinction is usually made between two physical processes:

1) forced convection, where heat transfer is due to an external force field and is effected for very high velocities of motion and relatively small temperature differences. In this case it is possible to disregard the influence of the thermal state of the gas on the velocity field and on the values of the transport coefficient;

2) free or natural convection, where the process of heat transfer takes place for very large temperature differences and arises as a result of density changes induced by the flux of heat, e.g., in the heating of a layer of gas or liquid from below, as is usually done in laboratory investigations. Here it is assumed that no causes other than thermal ones operate in producing the motion of the gas.

In nature, of course, neither process occurs in pure form. Usually causes arising from the influence of external forces, which we might term dynamical, and thermal causes, where the force causing the motion is Archimedes force or buoyancy, act simultaneously.

Observations have established that while no predominant periods of temperature and wind pulsation can be detected in the presence of relatively small super-adiabatic gradients, for larger super-adiabatic gradients wind and temperature pulsations on a much larger scale appear in the surface layer of the atmosphere.

Ascending motions of warm air alternate with slower downdrafts of cold air; the air masses involved in these motions are considerably larger than the elementary particles that participate in turbulent mixing, and the motions of the latter are merely superposed on the overall background of the alternation of ascending and descending air currents. In such cases we may speak of the process of thermal convection.

One can see such motion by observing the air jets quivering over strongly heated surfaces on a hot day (asphalt-surfaced roads, iron roofs, etc.), or the state of smoke clouds in the presence of steep temperature gradients in the surface layer. In the latter case individual puffs of smoke

("bubbles") are seen to burst upwards and may be traced to a considerable height above the top of the cloud.

Thermal convection develops most frequently over land. In the warm part of the year, when sufficiently well-developed, it causes the formation of cumuliform clouds; these have therefore been termed convective clouds.

However, the details of the development of thermal convection have not yet been studied fully. One would expect the development of thermal convection to be similar to the pattern observed in laboratory experiments on the regular convective circulation of air over strongly heated plates. The air layer above such plates breaks down into regularly arranged cells, with an ascending motion at the center of each cell and descending motion along the sides (Bénard cells). The theory of regular convection has been worked out in fairly great detail. In nature, on the other hand, where the formation of such regularly distributed cells is hampered by the uneven heating of the earth's surface, the development of the phenomenon might be visualized as follows. At first when the earth's surface is warmed by the sun there is an upward transfer of heat by conduction and emission. However, as the latter mechanisms do not suffice to remove all the heat received by the lower layer of air this layer becomes overheated and therefore lighter. A very small disturbance will then be sufficient for the overheated masses to break away in the form of an updraft, while cooler masses rush in from above and from the sides. The circulation thus formed gradually extends over greater layers of air.

Observations by A. A. Skvortsov at Tashkent show that thermal instability does not appear at once from the surface to the level of cumuliform cloud; immediately above the surface one finds a series of thinner layers or "levels" with definite physical properties between which a jump-like transfer takes place by means of air bursts (similar to those mentioned in the previous paragraph and to the bursts observed in smoke clouds).

Chapter 14

TEMPERATURE OF THE FREE ATMOSPHERE

§ 1. Mean temperature distribution at various levels in the troposphere

Air temperatures in the troposphere decrease regularly with height. This decrease is somewhat different at different latitudes, and therefore at any level above the earth's surface the mean temperature distribution differs from the sea-level one. Results of aerological observations of free-atmosphere temperatures may be represented in the form of isotherm charts of the various levels analogous to those for the earth's surface. A clearer and more convenient way of expressing the results is by relative topography charts (see Chapter 5) and north-south cross sections of the atmosphere; these provide a simultaneous picture of the distribution of temperature over both height and latitude.

We recall* that relative topography charts show twice the mean temperature of the atmospheric layer contained between two isobaric surfaces. An RT_{1000}^{500} chart, for instance, characterizes the mean temperature (according to height) of the lower troposphere between the 1000 and 500 mb levels.

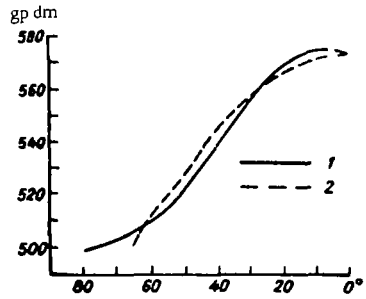


FIGURE 77. Mean latitudinal values of relative geopotential H_{1000}^{500} from equator to high latitudes. Winter

1—northern hemisphere; 2—southern hemisphere.

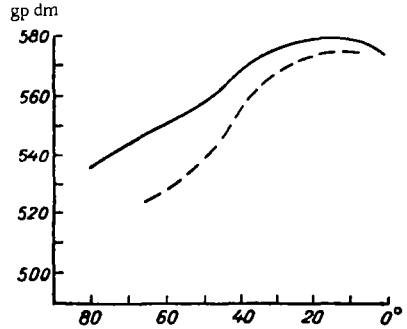


FIGURE 78. Mean latitudinal values of relative geopotential H_{1000}^{500} from equator to high latitudes. Summer

Figures 77 and 78 illustrate the mean latitudinal values of H_{1000}^{500} for winter (Figure 77) and summer (Figure 78) over the entire surface of the

* [Cf. p. 91.]

earth; Figure 79 gives a scheme showing the distribution of the mean temperature over the northern hemisphere in winter and summer.

From these figures one can conclude, firstly, that in the main layer of the troposphere (to a height of 5.5 km), as at the earth's surface, temperatures decrease from the equator polewards (the gradient dips in the direction of the pole); this is true in winter as well as in summer. However, the relationship changes in the tropopause layer and in the lower stratosphere, and the converse temperature pattern becomes established starting from a height of 12–14 km (that is, temperatures increase polewards).

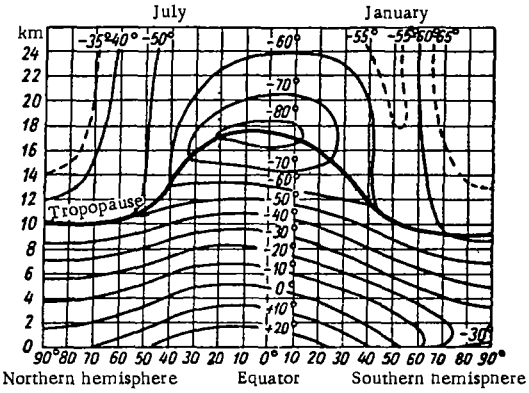


FIGURE 79. Vertical section across the atmospheric temperature field along the meridian

Thus the horizontal temperature gradient dips away from the equator toward the poles in the troposphere, then reverses direction in the lower stratosphere.

Horizontal gradients are particularly mild in the equatorial region (south of latitude 20° N. in winter and south of latitude 30° N. in summer), where they drop practically to zero in the upper troposphere. The steepest gradients occur between latitude 25 and latitude 40° N. in winter, and between latitude 35 and latitude 45° N. in summer; they decrease again to some extent at higher latitudes.

The influence of continents and oceans on the distribution of temperature is considerable at the earth's surface and perceptible in the lower half of the troposphere. In winter the areas of cold in the troposphere correspond approximately to the terrestrial poles of cold and in both winter and summer deep troughs of cold and heat are visible on the charts. Thus on the winter H^{500}_{1000} charts troughs of cold with closed contours are found over north-east Asia and North America. A somewhat fainter cold trough is observed over Eastern Europe. Between the cold troughs lie warm ridges (e.g., over Scandinavia and the Gulf Stream). In summer the picture changes and cold troughs, though somewhat fainter, are observed over cool ocean regions (Bering Sea, east of Newfoundland).

The isograms are particularly crowded at latitude 30–40°, as is clearly seen on the winter H^{500}_{1000} charts. The zone of crowded isograms is best developed in the region of cold troughs and is fainter in the region of warm ridges. This zone, which ranges over the entire hemisphere, is called the upper-air planetary frontal zone.

§ 2. Vertical variation of air temperature

The lapse of air temperature with height in the troposphere averages 0.6° per 100 m. As more detailed analysis of aerological data will show, this lapse rate does not remain constant with height, even on the average. Four layers can be singled out in the troposphere above the surface layer depending on the magnitude of the vertical temperature gradients: 1) lower (boundary) layer, extending to a height of 1–1.5 km, in which the average daily gradient amounts to only 0.3–0.4°/100 m; 2) middle-troposphere layer between 1.5 and 5–6 km, with a gradient of 0.5–0.6°/100 m; 3) upper-troposphere layer between 6 and 8–9 km, with a gradient increasing to 0.65–0.75°/100 m; and, finally, 4) the tropopause fluctuation layer, located at a height of 9–12 km, in which the mean value of the gradient drops sharply to 0.5–0.2°/100 m. Above this level, in the stratosphere, the temperature either remains nearly constant or rises slightly to a height of the order of 35 km.

The most unstable gradients are found in the lowermost boundary layer, where the influence of the heating and cooling of the earth's surface is strongly felt. Gradients here change over the year and depend on the time of day. The mean daily value (0.3–0.4°/100 m) is not representative. On summer days, when the earth's surface experiences strong heating, temperature gradients in the layer extending up to 300–500 m may reach super-adiabatic values. At night, when the surface is cooled, gradients in this layer often change sign. In winter inversions prevail even in daytime. Nighttime inversions can be very considerable — up to several degrees; super-adiabatic gradients in this layer, however, usually do not exceed 1.2–1.5°/100 m and seldom reach 2°/100 m. This is due to the intense vertical transfer which develops in daytime when the surface is strongly heated and super-adiabatic gradients are formed; this vertical transfer reduces the vertical temperature lapse rate and brings it closer to the adiabatic rate.

Above 1.5 km in the main layer of the troposphere mean vertical temperature gradients are fairly stable through the year and vary only slightly in middle-latitude regions. A certain amount of variation, depending on geographic conditions, is observed in the upper troposphere.

Above the upper boundary of the troposphere lies the fourth layer listed above, the tropopause. Its height and temperature depend largely on latitude, time of year and processes developing in the atmosphere; this question will be considered in greater detail later. Data for points at different latitudes are given in Figure 80a and 80b to illustrate the variation of temperature with altitude. Apart from the regularities already mentioned one can see that the temperature drop extends to a significantly greater height in low latitudes (up to the 90–100 mb level at the equator) than in high latitudes; correspondingly the lowest temperature (down to −80°) in the troposphere is observed near the equator, while the minimum in northern regions amounts to −50, −60° at heights of about 10 km. In the equatorial zone the temperature begins to increase after reaching the minimum while in the polar regions it continues to decrease slowly.

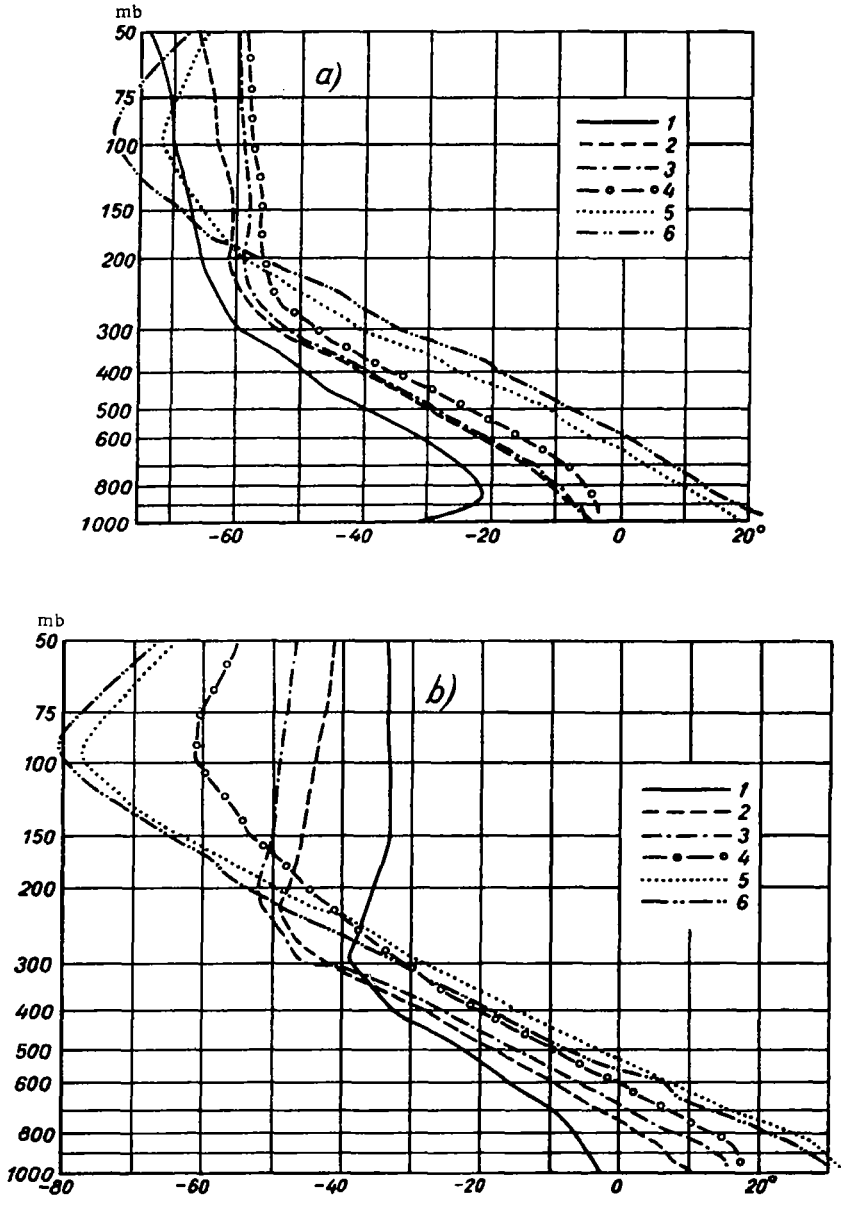


FIGURE 80. Vertical distribution of temperature at different latitudes. a) winter; b) summer

1- floating polar research station No. 4; 2-Murmansk; 3-Moscow; 4-Erevan-Tbilisi; 5-Bahrein; 6-Aden.

§ 3. Annual and daily march of air temperature in the troposphere

The periodic variations of air temperature (annual and daily) which are so pronounced in the boundary and especially in the surface layers continue to be felt throughout the troposphere; their character, however, changes somewhat with altitude.

Observations show that the amplitude of daily temperature oscillations decreases with height; this decrease, moreover, is much more rapid in winter than in summer. At the same time a phase shift of the daily wave takes place to a height of about 2 km; that is, the maxima and minima lag behind those at the surface This change in the character of the daily oscillations is easily explained by the basic mechanism of heat transport from th the earth's surface— turbulence. But in addition to this mechanism daily oscillations are also determined by the radiant influx of heat, which consists of the thermal emission of the earth's surface and atmosphere and also of the short-wave radiation of the sun. Starting from a certain height the radiant influx begins to dominate, and in the middle and upper troposphere (above 2 km) the daily temperature oscillations are determined mainly by the latter factor. At these altitudes the maximum occurs in daytime (13–14 hrs) and the minimum in the morning (5–7 hrs). Daily amplitudes in the free atmosphere are limited— 1–2° in winter and 3–4° in summer.

There are as yet few reliable data on the daily march of temperature in the stratosphere, though it is certain that daily oscillations do take place there; this can also be seen from Table 51.

TABLE 51

Amplitude of daily march of air temperature (Moscow)

Height (km)	2	4	6	8	10	12	14	16	18	20	22	24
Amplitude (°)	1.4	2.5	2.4	5.3	6.9	3.9	4.2	5.8	6.6	7.8	9.8	13.5

The annual temperature oscillations in the troposphere are very clear-cut; at first their amplitude decreases somewhat with altitude but then increases (in the upper troposphere), and a drop is observed only upon crossing into the stratosphere; later the amplitude increases again.

For illustration results of researches carried out in the Belmar area (latitude 40.2° N., longitude 74.1° W.) are given in Figure 81 for the four seasons. These data confirm what we said earlier and show that the amplitude of annual temperature oscillation in the stratosphere is not smaller than in the troposphere.

With increasing height there is a certain delay in the occurrence of the annual temperature minimum and (especially) maximum compared with the earth's surface. This lag increases with height, and in the middle troposphere the extrema are shifted to February and August respectively. In the upper troposphere any further lag is imperceptible while in the stratosphere the minimum and maximum occur even earlier than in the troposphere (June–July and December–January). Observing the seasonal distribution of temperature one finds that air temperatures in the troposphere are somewhat higher (by 2–3°) in autumn than in spring; in the lower

stratosphere, on the contrary, autumn temperatures are somewhat lower than spring ones.

This shift in the time of occurrence of the extrema in the stratosphere toward the solstices, and also the absence of phase shifts in the daily oscillations in the troposphere, certainly means that the thermal state of the

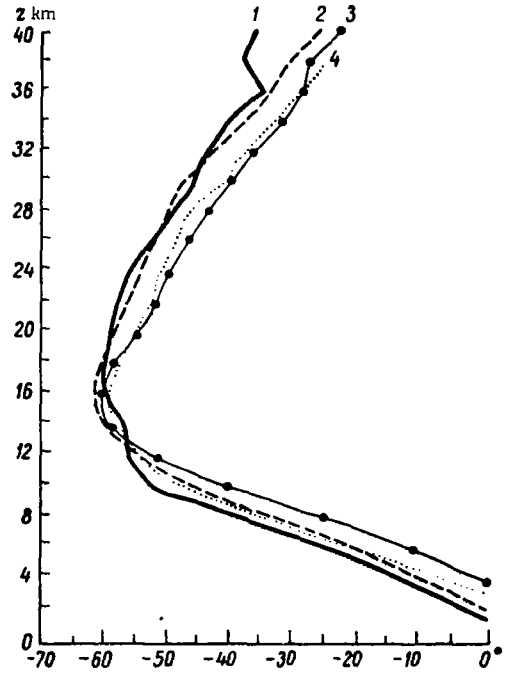


FIGURE 81. Vertical distribution of mean air temperature over Belmar

1-January-March; 2-April-June; 3-July-September; 4-October-December.

atmosphere at these altitudes is greatly influenced by processes of radiative heat transfer. This peculiarity of the annual march in the atmosphere can apparently be attributed to the fact that the lower stratosphere does not receive heat from the earth's surface alone, and that its temperature regime is largely regulated by changes in its content of ozone, which is a strong absorber of solar radiation.

§ 4. Height and temperature of the tropopause

A layer of transition to the stratosphere is permanently present at the top of the troposphere; this is the so-called tropopause. The height and temperature of the tropopause depend on latitude, time of year and processes developing in the atmosphere.

The temperature at the tropopause, as one can see from the data cited

in the previous section, is lowest in the equatorial regions. The lowest temperature recorded at the tropopause is about −90° (−91.9° has been reported for 5 December 1912 over Jakarta at a height of 17.0 km, and a temperature of −92° was recorded over Agra on 4 November 1928 at a height of 16.8 km).

The height of the tropopause is also different at different latitudes; on the whole the level of the tropopause rises from the poles towards the equator, and there is a simultaneous decrease in its temperature (see Figure 79).

However, more detailed study shows that the equatorward rise in the level of the tropopause is not really as smooth and regular as shown in Figure 79 for a schematized mean chart. A break in the even slope of the tropopause occurs at latitudes of about 30–40°, as one can see from Figures 13 and 14. This break is present in all months, so that there are virtually two tropopauses, a tropical and a polar one. The tropical tropopause is nost clear-cut at latitude 0–35° N. but can sometimes be traced to latitude 45° N. and partly overlaps (for a distance of 5–10° of latitude) the polar tropopause. At these latitudes (35–45°) two tropopauses lying one above the other are thus observed and the distance between them varies from about 2.5 to 5 km in different months (the distance is greater in winter).

As was already obvious from the data cited above, the height of the tropopause and temperature of the lower stratosphere at any point display a clear-cut annual march; the temperature of the stratosphere is higher in summer than in winter, and the height of the tropopause is greatest in late summer to early autumn and lowest in late winter or early spring. In middle latitudes the tropopause lies at a height of 11–12 km in summer and at a height of 9–10 km in winter. However, very large deviations from the mean position are observed in individual cases. At Moscow, for example, the extreme positions of the tropopause are 6–7 and 13–14 km, i.e., deviations of 6–7 km are possible. We note that changes in the level of the tropopause have been linked with pressure systems: a low tropopause is observed in cyclones and a high one in anticyclones. Furthermore, differences in temperature distribution are also observed for different pressure formations. For a normal surface pressure, the tropopause and lower stratosphere are isothermal; in the case of cyclones there is a sharp inversion in the tropopause layer and subsequent gentle temperature lapse in the lower stratosphere, and in anticyclones a moderate inversion occurs both in the tropopause layer and above it.

It has also been established that the tropopause rises when relatively warm air flows into an area and sinks when cold air advances.

As already pointed out, the higher the tropopause the lower (usually) its temperature. However, important exceptions to this rule are observed; in winter over the Arctic and Antarctic, for example, the tropopause is low and at the same time cold. A very cold tropopause is also found in the area of the poles of cold in the northern hemisphere.

Observations at permanent and floating stations in the Arctic, and also in the Antarctic, have made it possible to determine some of the characteristic features of these regions. In the Arctic it was found that the temperature of the atmosphere at heights of 20–25 km reached −70 to −80° during the polar night; that is, temperatures here are lower than at the same levels in the equatorial zone. In Antarctica even lower temperatures (down to −81.2°) have been reported at the same height in winter. At the

same time temperatures at these heights are comparatively high in the summer period, so that the amplitude of annual oscillations turns out to be very considerable.

§ 5. Inversions

Although on the whole the troposphere is characterized by a smooth lapse in temperature with height, considerable deviations from the mean are very often observed. Such deviations, which can occur at various heights and over layers of varying thickness, take different forms: the vertical lapse may slow down very considerably or vanish entirely (isothermy), or it may even be replaced by an increase in temperature with height (inversion). These layers are thermodynamically very stable and tend to hamper the development of ascending motions in the atmosphere; they have therefore been called restraining layers.

The conditions of formation and thermodynamical significance of these layers, though in principle the smae for all types, are best seen in the case of inversions.

We will use the term layer of temperature inversion, or simply inversion layer, to indicate an atmospheric layer in which the temperature increases with height (the lapse-rate $\gamma = -\frac{\partial T}{\partial z}$ is negative). The height of the lower boundary of the inversion is called the inversion height, the thickness of the inversion layer— the inversion depth, and the total temperature increment from the base to the top of the inversion layer— the magnitude of the inversion. Measurements show that these parameters can vary over a very wide range depending on the character of development and type of the inversion.

The following two types of inversion are distinguished according to height: 1) surface inversions, beginning directly at the earth's surface, and 2) free-atmosphere inversions, the height of which can vary considerably. Sometimes not one but two or more inversion layers are observed simultaneously.

The thickness of inversions varies from a few meters to 2–3 km; their magnitude also varies and can reach several degrees (up to 10° and even more).

Inversions are not a particularly rare phenomenon. It would be more correct to say that their presence in the atmosphere is the rule; for example, the presence of inversions was detected in 98 % of all observations during the 1934–1935 period at Moscow.

The processes responsible for the formation of inversions can be extremely varied, and in actual conditions different factors often act simultaneously. Classification of inversions according to origin is therefore very difficult. However, if we single out the main responsible factors it is possible to distinguish between several types of inversion. Let us consider the principal types.

Surface inversions include: 1) radiation inversions, 2) orographic inversions, 3) warm air inversions and 4) spring or snow inversions.

1. Radiation inversions are the most widespread. They form near the earth's surface when the latter is strongly cooled as a result of thermal emission. The air layers directly above the surface are also

cooled, and as a result the temperature increases instead of decreasing in a certain layer of air. Obviously such inversions develop in cases where the earth's surface gives off more heat by emission than it receives. This happens at night and, sometimes in winter, throughout the day as well. A distinction is therefore made between night inversions and winter inversions. Conditions which favor the development of radiation inversions are the absence of cloud (intensive emission), feeble movement of air (not exceeding 2–3 m/sec), and the presence of large temperature oscillations from day to night. Usually the thickness of the layer of radiation inversion varies from 10–15 to 200–300 m and, under very favorable conditions, to 300–400 m. The magnitude amounts to about half the daily amplitude at the height of the meteorological booth, i.e., several degrees. Such inversions are characteristic of all regions but especially of deserts.

Radiation inversions attain very large thicknesses and magnitudes in winter in Yakutia, where they are enhanced by other factors acting simultaneously and therefore have a somewhat more complex character. Their thickness can reach up to 3 km and their magnitude up to about 15°.

2. Orographic inversions are a form of radiation inversion in which the radiation effect is heightened by features of the orography. In a hilly locality, for example, cool air flows down from the uplying parts into hollows or valleys, where cooling is stronger even otherwise due to weaker transfer.

3. Warm air inversions occur when warm air masses are transported horizontally into an area (advection) and travel over a cooler underlying surface. The lowermost layers of the warm flow release part of their heat to the underlying surface and are thus cooled more effectively than the upper layers. Such inversions are especially frequent, for instance, when warm sea air is carried over cool land surfaces in winter.

4. Spring or snow inversions are also related to the advection of warm air, in this case over a surface covered by snow. Their development is mainly due to the expenditure of large amounts of heat (released by the air) in the melting of snow. Their thickness and magnitude are usually limited.

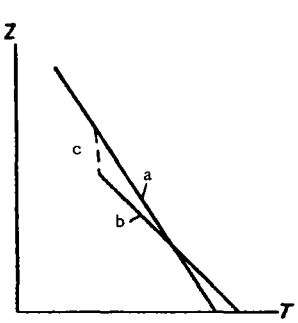


FIGURE 82. Formation of frictional inversion

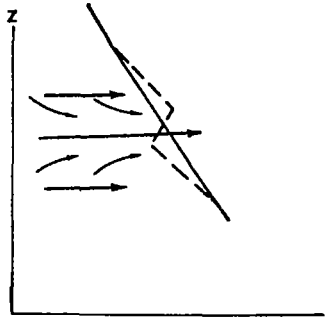


FIGURE 83. Formation of dynamical inversion

Free-atmosphere inversions include: 1) frictional inversions, 2) dynamical inversions, 3) anticyclonal or compressional inversions and 4) frontal inversions.

1. Frictional inversions are transitional from surface to free-atmosphere inversions. They form at a height of a few hundred meters above the surface, at the upper limit of the friction layer, i.e., the atmospheric layer in which the direct influence of the ground is particularly strong and turbulent transfer particularly intesive. Qualitatively their formation can be explained as follows. A vertical gradient close to the adiabatic is formed in the friction layer as a result of turbulent mixing, and the stratification curve moves from position a to position b (Figure 82). The temperature drop in the upper part of the layer leads to a discontinuity at its boundary and to the appearance of an inversion (c). The vertical thickness of such inversions is limited to a few tens of meters.

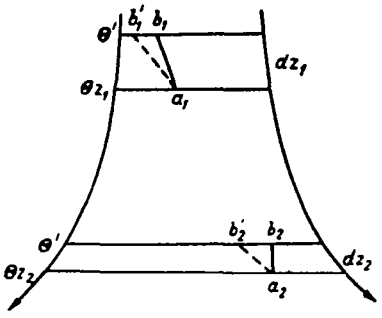


FIGURE 84. Formation of compressional inversions

2. Dynamical inversions develop in the free atmosphere in layers with large wind velocities. The formation of such inversions is as follows. A rapidly moving air flow sucks in air at lower velocities from adjoining layers (Figure 83), descending motions are created at the top of the high-velocity layer and ascending motions at the bottom. The temperature of the air particles rises adiabatically in the zone of descending motions and drops adiabatically in the zone of ascending motions. As a result a redistribution of temperature takes place. The original stratification line a is replaced by a second, more complicated, stratification line b; an inversion is formed in the central section of the high-velocity layer. In the layers adjoining the latter vertical temperature gradients increase to the adiabatic level.

3. Compressional inversions, or inversions of subsidence, are the most interesting; they are also called anticyclonal, since they usually develop in anticyclones. In such pressure regions one finds, as a rule, a general descending motion in the middle troposphere and an outflow of air from the center toward the edges in the lower troposphere. In this process a certain layer of air of thickness dz_1 occurring at the level z_1 drops to the level z_2 and is compressed to the thickness dz_2. Its cross section, equal to s_1 in the initial position, increases to s_2 due to spreading. As a result of the compression of the layer a change takes place in its stratification.

Let us denote by p_1, T_1, ρ_1 and γ_1 the pressure, temperature, density and vertical temperature gradient* at the initial position of the layer (level z_1), and by p_2, T_2, ρ_2 and γ_2 the same characteristics at the final position at the level z_2 (Figure 84).

We will assume that the layer consists of the same particles at all times, that the vertical displacements of the layer are adiabatic, and therefore that the potential temperature Θ remains constant for individual particles. Since Θ changes with height and is different at the initial and final positions

* [Actually the lapse rate, or the negative of the temperature gradient $\frac{\partial T}{\partial Z}$. Z is measured upwards. Cf. the author's definitions, Ch.6, §4.]

of the layer under consideration, one has $\frac{d\theta}{dz_1} \neq \frac{d\theta}{dz_2}$, whence

$$\frac{d\theta}{dz_2} = \frac{d\theta}{dz_1}\frac{dz_1}{dz_2}. \tag{1}$$

But since $s_1\rho_1 dz_1 = s_2\rho_2 dz_2$, we have

$$\frac{dz_1}{dz_2} = \frac{s_2\rho_2}{s_1\rho_1} = \frac{s_2T_1\rho_2}{s_1T_2\rho_1}, \tag{2}$$

therefore

$$\frac{d\theta}{dz_2} = \frac{d\theta}{dz_1}\frac{s_2\rho_2T_1}{s_1\rho_1T_2}. \tag{3}$$

However, it was shown earlier that

$$\frac{d\theta}{dz} = \frac{\theta}{T}(\gamma_a - \gamma).$$

Consequently, instead of (3) we write

$$\frac{\theta}{T_2}(\gamma_a - \gamma_2) = \frac{\theta}{T_1}(\gamma_a - \gamma_1)\frac{s_2\rho_2T_1}{s_1\rho_1T_2}, \tag{4}$$

which gives

$$\gamma_a - \gamma_2 = \frac{s_2\rho_2}{s_1\rho_1}(\gamma_a - \gamma_1), \tag{5}$$

or

$$\gamma_2 - \gamma_1 = (\gamma_a - \gamma_1)\left(1 - \frac{s_2\rho_2}{s_1\rho_1}\right). \tag{6}$$

To analyze the resulting formula, let us consider two cases:

1) the initial state of the atmosphere is unstable and therefore $(\gamma_a - \gamma_1) < 0$; then upon compression $\left(1 - \frac{s_2\rho_2}{s_1\rho_1}\right) < 0$, and therefore $(\gamma_2 - \gamma_1) > 0$, i.e., the temperature gradient increases, while remaining positive. Upon expansion of the layer $\left(1 - \frac{s_2\rho_2}{s_1\rho_1}\right) > 0$, therefore $(\gamma_2 - \gamma_1) < 0$, i.e., the temperature gradient decreases. It may fall to zero and even become negative.*

2) the initial state of the atmosphere is stable, i.e., $(\gamma_a - \gamma_1) > 0$; then upon compression $\left(1 - \frac{s_2\rho_2}{s_1\rho_1}\right) < 0$ and therefore $(\gamma_2 - \gamma_1) < 0$, i.e., the temperature gradient decreases and the state of the atmosphere becomes more stable. Upon expansion $\left(1 - \frac{s_2\rho_2}{s_1\rho_1}\right) > 0$, so that $(\gamma_2 - \gamma_1) \gtrless 0$, i.e., the temperature gradient increases, the state of the atmosphere becomes less stable and may even become unstable.

Thus, on compression, an initially unstable state becomes even more unstable and an initially stable state even more stable. On expansion the converse change takes place.

The development of inversions is possible in the case of compression for a stable initial state. In this case $(\gamma_2 - \gamma_1) < 0$ or $\gamma_2 < \gamma_1$ and the stability of the stratification increases when the layer sinks downwards. In particular γ_2 may be equal to zero, which happens when

$$\frac{s_2\rho_2}{s_1\rho_1} = \frac{\gamma_a}{\gamma_a - \gamma_1}.$$

Consequently, in this case isothermy is established in the final state.

If, on the other hand,

$$\frac{s_2 p_2}{s_1 p_1} > \frac{\gamma_a}{\gamma_a - \gamma_1},$$

then $\gamma_2 < 0$ and an inversion is formed.

O. P. Petrenchuk has shown that stable layers gradually transform into inversions in developing anticyclones; when the descending motions cease the inversions fade away.

Compressional inversions are characterized by great vertical and horizontal extension. They develop most often at heights of 1–2 km, and in winter sometimes at lower levels. Their appearance is associated not only with a rise in temperature but also with drying of the descending air; relative humidity in the inversion layer sometimes drops to 20–30%. Layers of haze frequently appear under the inversion, while in the cold season fog and clouds are often formed.

If clouds are present in the initial state the process will be far more complex.

4. Frontal inversions appear in frontal zones when warm air rises over the wedge of cold air. In this process the isotherms in the warm

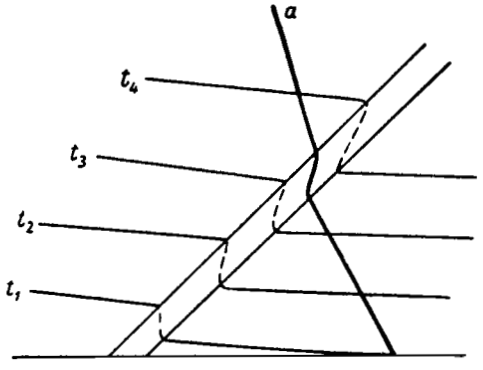


FIGURE 85. Frontal inversion

air mass lie at a higher level than those in the cold air mass and a discontinuity appears in the transitional zone (Figure 85). The discontinuity will depend on the temperature contrast between the air masses, and one of a veriety of vertical gradients may appear in the transitional zone: inversion and isothermal gradients in the case of large contrasts, weakly positive gradients for small contrasts. In complex fronts the inversion may break down into several layers, and a complex thermal structure may be observed in the atmosphere in the frontal zone.

§ 6. Theoretical considerations on the vertical distribution of temperature in the free atmosphere

The discussion of the temperature distribution given in the previous sections was based on observational data. In addition to giving a qualitative interpretation of the stated regularities it is desirable to investigate them theoretically and to obtain quantitative relations.

The earliest theories of the vertical distribution of temperature in the atmosphere were based on the assumption that radiative equilibrium prevailed in the atmosphere. In other words, the amounts of absorbed and emitted radiation should be the same in any layer of the atmosphere.

We will consider one of the elementary variants of the theory of radiative equilibrium in the atmosphere, using the simplified equations of transport for radiant energy obtained in Chapter 10, § 2.

Let us assume that the coefficient of absorption of long-wave radiation in the atmosphere does not depend on the wavelength. The equations of radiative transport then assume the following form:

$$\frac{dG}{dm} = k(G - E),$$

$$\frac{dU}{dm} = k(E - U). \tag{7}$$

Adding and subtracting these equations we obtain, subject to the condition of radiative equilibrium $G + U = 2E$,

$$\frac{d}{dm}(G + U) = k(G - U),$$

$$\frac{d}{dm}(G - U) = k(G + U - 2E) = 0. \tag{8}$$

Thus from the above we have

$$G - U = \text{const}$$

or

$$U - G = F_r = \text{const}, \tag{9}$$

i.e., in radiative equilibrium the effective flux of long-wave radiation F_r remains constant with height; this, as we saw, is not observed in the troposphere.

Using relation (9) we obtain, from equation (8),

$$\frac{d}{dm}(G + U) = -kF_r.$$

From this we find

$$G + U = kF_r m + C.$$

Since for $m = m_\infty$ (at the top of the atmosphere) we have the obvious boundary condition $G = 0$, recalling (9) we obtain

$$C = F_r(1 + km_\infty).$$

Thus we can write

$$G + U = F_r[1 + k(m_\infty - m)]$$

or, bearing in mind the condition of radiative equilibrium $G + U = 2E = 2\sigma T^4$,

we find that

$$E = \sigma T^4 = \frac{1}{2} F_r [1 + k (m_\infty - m)]. \quad (10)$$

Since in the stratosphere $m = m_\infty$, with sufficient accuracy

$$\sigma T^4 = \frac{1}{2} F_r, \quad (11)$$

where F_r is the long-wave emission of the earth and atmosphere escaping into inter-planetary space.

Obviously, on the global average this quantity is equal to the quantity of solar energy incident (with allowance for the albedo) on one square centimeter of the earth's surface. This quantity of solar energy, which we denote by S, can easily be calculated from the following formula:

$$S = S_0 \frac{\pi R^2}{4 \pi R^2} (1 - A) = \frac{S_0}{4} (1 - A),$$

where S_0 is the solar constant, R the radius of the earth, and A the mean value of the earth's albedo.

Thus $F_r = S$. Consequently,

$$\sigma T^4 = \frac{1}{2} F_r = \frac{1}{2} S. \quad (12)$$

The above relation makes it possible to arrive at a rough estimate of the temperature of the stratosphere.

The physical meaning of the result obtained according to the theory of radiative equilibrium can be clarified with the aid of the following model.

Let us consider a body lying between two infinite parallel planes at a temperature T_1. We will assume that the body in question is in radiative equilibrium with these planes. Obviously, the temperature of the body in this case is also T_1. If one now removes one of the planes, the situation of the body becomes analogous to that of the stratosphere. The temperature of the body changes when one of the planes is removed, and if radiative equilibrium is maintained this temperature T_2 can be determined from the following equation:

$$2 \sigma T_2^4 = \sigma T_1^4,$$

i. e.,

$$\sigma T_2^4 = \frac{1}{2} \sigma T_1^4,$$

where σT_1^4 is analogous to the quantities F_r or S.

Assuming, for example, that $T_1 = 259°$, we obtain $T_2 = 218°$, which is in close agreement with observations.

However, although calculations performed according to the latter formula give correct values of the stratospheric temperature, they are in poor agreement with the actually observed distribution of temperature in the troposphere. Attempts to eliminate these divergences by introducing corrections into the theory of radiative equilibrium have not led to good results. This shows that the theory of radiative equilibrium must be revised to include other processes influencing air temperature. Even aside from the fact that it does not make a sufficiently correct allowance for the absorption of radiation in the atmosphere, the theory disregards the influx of heat due

to adiabatic changes of pressure, turbulent transfer, advection and phase transformations of moisture in the atmosphere. Obviously, for a general solution of the problem it is necessary to allow for all these factors and to take the equation of heat influx in the atmosphere in its more complete form

$$c_p \rho \frac{dT}{dt} = A \frac{dp}{dt} + \varepsilon_r + \varepsilon_e + \varepsilon_{ad} + \varepsilon_{ph} \qquad (13)$$

(all symbols are as before).

To (13) we must add the equations which give the individual fluxes; furthermore, we have to know the numerical values of a number of parameters appearing in these equations. Obviously, the solution of such a system of equations presents very great difficulties and would be impossible without certain simplifications. I. A. Kibel' (1943) gave the first approximate solution of (13); he assumed for simplicity that

$$\frac{dT}{dt} = 0, \quad \frac{dp}{dt} = 0, \quad \varepsilon_{ph} = 0, \quad \varepsilon_{ad} = 0.$$

Thus Kibel' reduced equation (13) to the form

$$\varepsilon_r + \varepsilon_e = 0, \qquad (14)$$

i. e., he solved the problem for the mean annual distribution of temperature $\left(\frac{\partial T}{\partial t} = 0\right)$ on the assumption that the thermal regime of the atmosphere is dependent only on the radiative and turbulent transport of heat.

Employing the simplified expressions for ε_r and ε_e given earlier, Kibel' confined his attention to vertical mixing, giving these expressions the form

$$\varepsilon_r = k\rho_{vap}(U + G) + k'\rho_{vap} S - 2k\rho_{vap} fE, \qquad (15)$$

$$\varepsilon_e = \bar{c}_p \rho \frac{d}{dz}\left(k_z \frac{dT}{dz}\right), \qquad (16)$$

where k and k' are the coefficients of absorption of short- and long-wave radiation and f is a coefficient indicating the departure from black-body emission.

Insertion of these expressions into equation (14) makes it possible to solve it for $E = f\sigma T^4$ and then obtain the value of T. Specific corrections were introduced into Kibel's solution by various investigators. Thus allowances were made for the diffuseness of the propagation of long-wave radiation, for the frequency dependence of the absorption coefficient and for the variation of the coefficient of turbulence with height.

The calculations produced values which were closer to actual data in the troposphere; however, for the lower stratosphere the calculated temperatures proved to be substantially different from the experimental ones.

The next step in the development of the theory was due to E. N. Blinova, who considered the large-scale horizontal transport. The calculation of this essential factor gave Blinova the possibility of computing the vertical temperature distribution as a function of latitude for a pure zonal circulation. This work was refined by various authors in later years. As a result of all these works more or less accurate allowances were made for the frequency dependence of absorption and diffuseness of long- and short-wave propagation, condensation of water vapor in the atmosphere, and evaporation from the earth's surface; the latitude dependence of such parameters

as the albedo, optical mass of the atmosphere, and amount of evaporated and condensed moisture was also taken into account.

Without going into the mathematical aspects of the solution or the choice of necessary parameters characterizing the given phenomenon, we present the vertical distribution of the mean annual temperature at various latitudes, as obtained by theoretical calculation, in Figure 86. As one can see from the figure, the theoretical calculations reflect the character of the north-south distribution of temperature correctly and show all the main features, e.g., the temperature difference between the north and south poles and the shifting of the thermal equator into the northern hemisphere. The divergence between the calculated and observed temperatures is small— no more than 2° in the northern hemisphere and somewhat higher in the southern hemisphere.

The vertical distribution of temperature in the troposphere as calculated theoretically (Figure 86) is also in fairly satisfactory agreement with observed data.

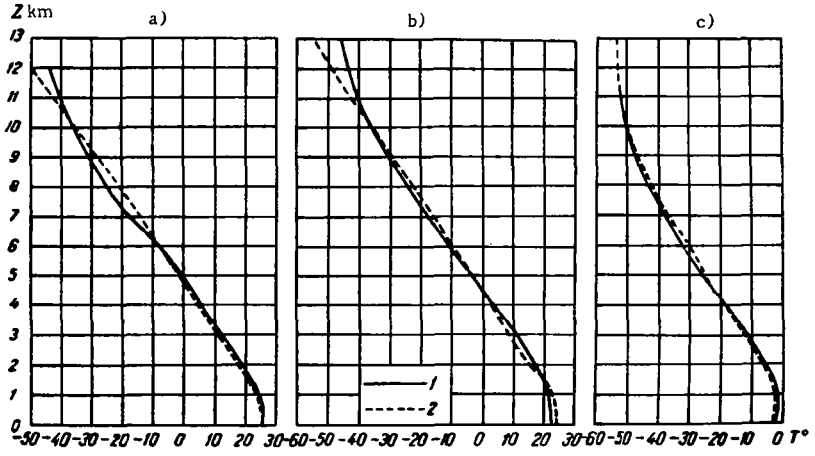


FIGURE 86. Vertical distribution of temperature

a—at equator; b—at 30° N. lat; c—at 60° N. lat; 1—calculated; 2—observed.

Thus for the most part the problem of the mean annual distribution of temperature in the atmosphere has now been solved theoretically up to heights of the order of 15 km. The next problem, which is even more difficult, is to carry out a theoretical study of the variation of temperature with time.

As to the thermal regime of higher atmospheric layers, a theoretical treatment of the question has only lately become possible thanks to experimental data obtained in recent years. The stratosphere and mesosphere, as was pointed out in Chapter 4, are characterized by the presence of a temperature maximum at a height of about 50 km and minimum near 80 km, and also by the presence at all heights of a clear-cut daily march of temperature.

An analysis of theoretical studies relating to the temperature regime of the upper atmospheric layers (due to K. Ya. Kondrat'ev and O.P. Filipovich)

shows that in these layers the atmosphere, once again, does not occur in a state of radiative equilibrium. General chilling is observed in the lower half of the stratosphere and radiative heating in the upper half. The presence of the temperature maximum and minimum — at heights of 50 and 80 km respectively — is doubtless to be attributed to the presence of a maximum and minimum of heat influx at the corresponding levels. The reason for this, and also for the daily temperature march, is the vertical and hourly variation of the absorption of solar ultraviolet by ozone. This is also indicated by the fact that the thermal regime of the mesosphere is directly influenced by variations in solar activity. Although the thermal regime of the upper atmospheric layers is governed primarily by radiation, the question of the role of other factors and particularly turbulent mixing still remains open.

The supposition that the atmosphere is in thermal equilibrium at heights greater than 1000 km has not been confirmed. The experimental data available are still too scanty for ascertaining the causes responsible for the heating of the upper atmospheric layers. Further accumulation of reliable experimental data is required.

Chapter 15

THERMAL BALANCE OF THE EARTH'S SURFACE AND ATMOSPHERE

The solar energy incident on the earth's surface is transformed and propagated into the interior layers of soil or water, and also into the atmosphere, by the fluxes of heat considered earlier. The study of this redistribution and of all transformations of solar energy is assisted by setting up the thermal balance of the earth's surface and atmosphere which expresses the income and expenditure of heat in the earth-atmosphere system. The equations for the thermal balance are particular forms of the general law of conservation of energy and express the equality between the influx and expenditure of heat. These equations, which are valid for every instant in time, can be extended to time intervals (hour, day, year) by replacing the fluxes of heat by their sums over the period under consideration.

All these questions will be considered in the present chapter.

§ 1. Equation of thermal balance for the earth's surface

The earth's surface receives a considerable amount of energy from the sun. At the same time, heat is transferred from the surface or active layer to underlying soil layers and to the atmosphere, and energy is expended in the evaporation of water and in biological and other processes.

For land surfaces the equation of thermal balance should reflect the distribution of the heat which reaches the surface and is used in warming the soil and air and in evaporating moisture from the soil surface. In order to set it up it is necessary to sum up all the fluxes of thermal energy between the element of surface and surrounding space.

The first and most important component of the thermal balance is the radiation balance R, considered in Chapter 10. The second component of the thermal balance is the heat transfer in the soil, B, considered in Chapter 11. The third component is the heat transfer between the underlying surface and atmosphere, P. Finally, the fourth component characterizes the heat expended in evaporating moisture (or released in condensation) from the underlying surface, LE.

Taking the sum of all the heat fluxes between a unit area of the underlying surface and the surrounding space (i.e., soil and air), we obtain the following equation for the thermal balance of the underlying surface:

$$R = B + P + LE. \tag{1}$$

In the above the convention is that the radiation balance is positive for

an inflow of heat to the underlying surface. All the other fluxes, on the other hand, are positive for an outflow of heat (Figure 87) from the surface.

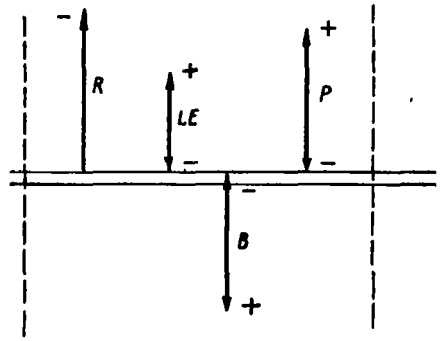


FIGURE 87. Thermal balance of land surfaces (schematic)

Since equation (1) contains only the four most important components of the heat balance for the underlying surface it is, of course, approximate. It does not include, for instance, such secondary terms as the heat transfer (positive or negative) due to precipitation penetrating into the soil, the heat expended in the photosynthesis of organic substances and several others. Also excluded from equation (1) is another, much more important, component of the thermal balance — the heat used up in melting snow (ice) or released in freezing water. This component is extremely important during the thaw period.

However, analysis of equation (1) makes it possible to elucidate many of the features of the thermal regime. Secondary factors can be taken into account when considering particular cases.

Equation (1) can be simplified even further by considering the thermal balance on the yearly average. Since the thermal state of the soil does not change on the yearly average, the heat transfer in the soil can be considered as zero ($B = 0$) over this period. In this case equation (1) has the form

$$R = P + LE. \tag{2}$$

For the surface of a dry soil, from which no evaporation occurs (i. e., $LE = 0$), we can write instead of (1)

$$R = B + P, \tag{3}$$

and on the yearly average

$$R = P, \tag{4}$$

i. e., all the radiant heat is exchanged with the atmosphere.

For the surface of oceans and seas, which occupy three-quarters of the entire surface of the globe, the equation of the thermal balance can be written by analogy with (1) in the form

$$R_0 = B_0 + P + LE, \tag{5}$$

where R_0 is the radiation balance of the water surface.

Heat transfer in the sea, B_0, has a slightly different meaning from heat transfer in the soil.

The point is that on land, owing to the smallness of horizontal temperature gradients in the soil, one can assume that only vertical transfer is present; this is determined mainly by the colorimetric conductivity of the soil. In the ocean, on the other hand, heat is transported not only by vertical exchange with the underlying layer but also by horizontal exchange with the surrounding body of water, the latter being largely controlled by ocean currents. The heat supplied (or removed) laterally is usually called the heat of water transport, or oceanic advection. Designating it by B_{ad}, one can write $B_0 = B + B_{ad}$, where B is that vertical transport which would be present in the absence of horizontal transfer.

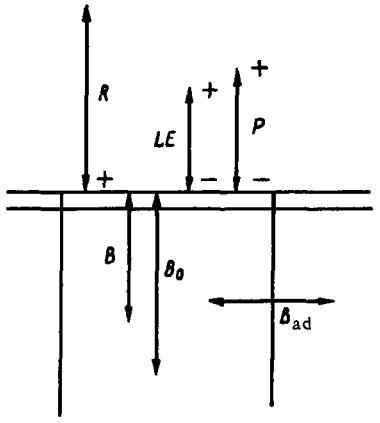


FIGURE 88. Thermal balance of the ocean surface (schematic)

Figure 88 provides a schematic representation of the thermal balance of the ocean. On the yearly average one can assume, as previously, that $B = 0$. Then $B_0 = B_{ad}$ and equation (5) for the mean annual ocean balance takes the form

$$R_0 = B_{ad} + P + LE. \qquad (6)$$

Equation (6), like (1), is approximate, since it fails to account for such important factors as the heat released (or absorbed) in the freezing of water (or melting of ice), or the heat delivered by run-off from land.

Considering the heat exchange between the underlying surface and atmosphere, we note that the inclusion of an advection term is not required in this case. Indeed, analysis of this question shows that advection does not directly influence either the thermal balance of the earth's surface or the thermal regime of the surface layer of air. Therefore in the earlier equations for the thermal balance of land surfaces one should set $P = P_e$ everywhere and assume that the exchange of heat between the earth s surface and atmosphere is entirely determined by the turbulent flux. From this it should not be inferred that horizontal advection in the atmosphere has no influence at all on the thermal balance of the surface. Such an influence is certainly present but it is manifested only indirectly, in terms of the changes that take place due to advection in all the components of the thermal balance — the radiant flux, the expenditure of heat in evaporation, the turbulent heat flux, and so forth.

§ 2. Equation of thermal balance for the atmosphere and earth-atmosphere system

In order to obtain the equation of thermal balance for the atmosphere one must obviously take the sum of all fluxes of heat in a vertical column of air of unit transverse section (1 cm^2) extending from the earth's surface (but excluding the latter) to the upper limits of the atmosphere. By analogy with the preceding discussion we have the following principal components of the balance: the radiation balance of the atmosphere R_A, turbulent heat

transfer P_e, the heat released in the condensation of water vapor L and, in the given case, the horizontal advection of heat across the walls of the atmospheric column under consideration P_{ad}.

For the radiation balance of the atmosphere R_A, which should be negative, we have the relation (cf. Chapter 10, equation (24))

$$R_A = F_0 - F_\infty + q', \tag{7}$$

where F_0 and F_∞ are the effective flux of long-wave radiation at the earth's surface and top of the atmosphere, respectively, and q is the direct and scattered radiation absorbed by the atmospheric column.

The amount of heat released in the condensation of water vapor can be taken to be Lr, where L is the latent heat of vaporization and the amount of moisture that has condensed and settled to the ground (i.e., precipitation).

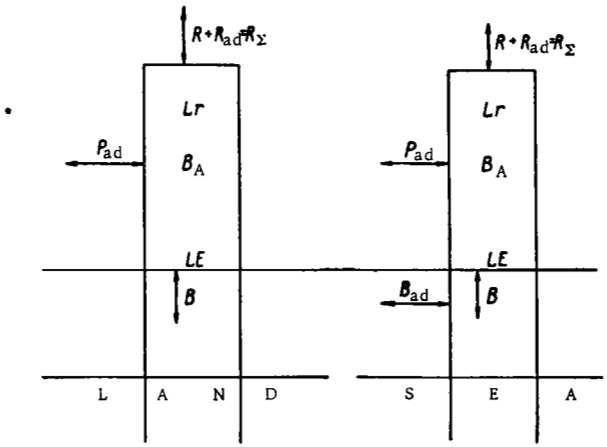


FIGURE 89. Thermal balance of the earth-atmosphere system (schematic)

Thus, remembering the convention regarding the signs of the fluxes, we write the equation of the thermal balance for the atmosphere as

$$-R_A = P_e + Lr - P_{ad} - B_A. \tag{8}$$

The term B_A accounts for the change in the heat content of the entire column of air over the period of summation. Over a year one can set $B_A = 0$, and the equation for the annual balance of the atmosphere then has the form

$$-R_A = P_e + Lr - P_{ad}. \tag{8'}$$

We note that on the yearly average the advection P_{ad} results in a transfer of heat from low to high latitudes. It is therefore associated with a heat loss in low latitudes and heat inflow in high latitudes.

To derive the equation of thermal balance for the earth-atmosphere system the vertical column of air should be continued through the land or sea layers down to the level at which temperature oscillations due to meteorological factors are absent. One must then take the sum of all fluxes of heat for this column.

Obviously, the required equation of the thermal balance for the earth-atmosphere system can also be arrived at by combining the equations of the thermal balance for the earth's surface and atmosphere.

Thus, adding (1) and (8) we obtain the following expression for an area over land

$$R_\Sigma = R + R_A = B + L(E - r) + P_{ad} + B_A = P_{ad} + L(E - r) + B_\Sigma, \quad (9)$$

where $B_\Sigma = B + B_A$ is the change in the heat content throughout the layer during the entire period of summation.

On the annual average $B_\Sigma = 0$, and

$$R_\Sigma = P_{ad} + L(E - r). \quad (10)$$

Over ocean surfaces one must also allow for the oceanic advection B_{ad} and the equation of the balance becomes

$$R_\Sigma = P_{ad} + L(E - r) + B_\Sigma + B_{ad}, \quad (11)$$

and on the annual average

$$R_\Sigma = P_{ad} + L(E - r) + B_{ad} \quad (12)$$

What we have been saying is illustrated schematically by Figure 89.

Considering the entire globe one may assume, on the yearly average, that $E = r$, i.e., that the amount of water evaporated from the earth's surface equals the amount of moisture settling out of the atmosphere in the form of precipitation. Furthermore, the horizontal transfer of heat in the atmosphere and oceans for the entire globe is obviously also zero. In this case the equation for the thermal balance of the entire globe becomes very simple, namely

$$R_\Sigma = (Q + q)(1 - A) + q' - F_{\text{eff}} = 0. \quad (13)$$

§ 3. Relative magnitudes of the components of thermal balance. Daily and annual march

To study the thermal balance one must either measure it or determine all its individual components. The direct measurement of the components is, however, a very complicated and tedious task due to the lack of simple and at the same time reliable apparatus. The methods evolved for measuring and determining the radiation balance — the principal component of the thermal balance — are the most refined; they involve actinometric observation as well as calculations, which were considered in Chapter 10.

Chapter 11 dealt with the state of the problem of determination of heat transfer between the surface and underlying layers of soil. The loss of heat in evaporation from the surface is given by the product of the latent heat of evaporation and the mass evaporated and is thus closely related to the question of determination of the amount of evaporation from the surface; this question will be considered further on (Chapter 17). Its determination is a fairly complicated problem. However, the most serious difficulties arise in the determination of the turbulent flux of heat; no reliable and convenient instruments are yet in existence for the direct measurement of this quantity. Thus the number of instrumental measurements of the

principal components of the thermal balance is very limited and is not sufficient for making more or less broad generalizations. In view of this situation methods for the indirect calculation of all these quantities were devised long ago, and have been very extensively developed in recent years.

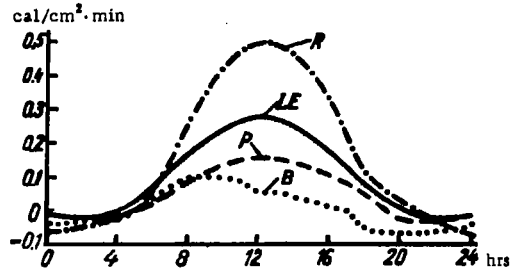


FIGURE 90. Daily march of the components of thermal balance in the Leningrad area. July

The methods in question are based on the utilization of data obtained by comparatively simple observation of the principal meteorological elements. They permit determination of individual components of the balance as averages over a fairly long interval (month, year) but are still too rough to allow calculation of these components over short intervals. Descriptions of the methods will be found in specialized instruction manuals; we will not dwell on them at any further length.

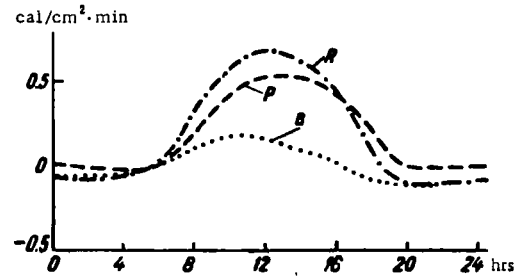


FIGURE 91. Daily march of the components of thermal balance in a semi-desert. July

The observed and calculated data now available for the thermal balance make it possible to draw a number of fairly reliable conclusions concerning the relative magnitudes of its individual components, both for individual geographic regions and globally.

We will consider firstly the daily march of the components of the thermal balance of the underlying surface.

The daily march of these components is shown in Figure 90 for the Leningrad area, and in Figure 91 for the semi-desert areas of Central Asia (Pakhta-Aral); the curves are for July. Above all these figures bear witness to the first-rank importance of the radiation balance. The value of

the radiation balance is greater than that of any other component of the thermal balance almost throughout the day. Other terms (heat transfer in the soil, B, and the turbulent flux of heat P_e) assume a dominant position in the balance only for brief periods near sunrise and sunset, at times when the radiation balance is changing sign.

According to calculations, at night the negative radiation balance is comparatively small and is compensated by the influx of heat from the soil and atmosphere to the underlying surface. In daytime the positive radiation balance is compensated by the sum of all the other components of the thermal balance. In humid climates (Leningrad) the loss of heat by evaporation is appreciably greater than the turbulent heat transfer, while in semi-desert conditions (Pakhta-Aral) this loss is zero in the month under consideration.

Extensive research along these lines has been carried out at the Main Geophysical Observatory under the direction of M. I. Budyko.

The heat transfer in the soil B is appreciably smaller than the loss of heat by turbulent transfer. From Figures 90 and 91 it is evident that most of the radiant heat received by the surface either enters the atmosphere directly, by turbulent transfer P_e (at Pakhta-Aral), or is expended in evaporation (Leningrad).

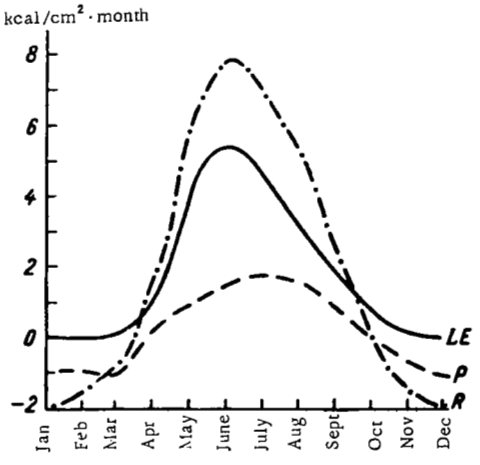


FIGURE 92. Annual march of the components of thermal balance over land. Barnaul

Of course, the correlations between components of the surface balance as represented in these figures are not universal. However, the essential features of the thermal balance as they appear from the examples here considered can be regarded as fairly typical. Thus the maximum of the thermal balance, as that of most of its components, occurs around noon, the only exception being the maximum of heat transfer in the soil, which is observed in the morning hours. The maximum amplitudes of the daily march of the components are recorded in summer and minimum amplitudes in winter.

The annual march varies very considerably with the particular climatic zone. In all regions, however, the dominant element in the annual march of the components is, once again, the radiation balance. The heat which

is received when the radiation balance is positive is used up chiefly in evaporation and turbulent transfer.

The least important term in the annual march is heat transfer in the soil; it attains maximum values, amounting to 4–12% of the radiation balance, in spring and autumn (Figure 92).

It should be mentioned that the annual march of individual terms in the balance is usually substantially different over land and over the oceans. Whereas the annual march of the radiation balance over oceans is generally analogous to that over land, turbulent heat transfer over oceans differs from that over land. This difference is chiefly due to differences in the processes of absorption of radiation and propagation of heat between the lower layers of land and sea.

The mean values of the annual sums of the components of the thermal balance over the surface of the globe are listed according to latitude zones, in kcal/cm^2· annum, in Table 52 (after M. I. Budyko).

The following inferences can be drawn from this table.

TABLE 52

Mean values of components of surface thermal balance, as functions of latitude, in kcal/cm^2· annum (after M.I. Budyko)

Latitude	Sea					Land				Earth				
	$\Sigma(Q+q)$	R	LE	P	B_{ad}	$\Sigma(Q+q)$	R	LE	P	$\Sigma(Q+q)$	R	LE	P	B_{ad}
60—50°N	88	34	34	18	—18	93	23	19	4	91	28	25	10	—7
50—40	109	54	51	15	—12	119	38	22	16	114	46	36	15	—5
40—30	136	78	73	12	—7	159	56	26	30	146	69	53	20	—4
30—20	151	100	85	7	8	184	64	23	41	163	86	60	20	6
20—10	156	110	89	5	16	182	74	36	38	163	101	75	14	12
10—0	149	107	76	5	26	149	79	58	21	149	101	72	9	20
0—10°S	152	107	81	7	19	143	75	59	16	150	99	76	9	14
10—20	155	107	97	9	1	161	69	44	25	156	99	85	13	1
20—30	147	94	87	10	—3	169	62	29	33	152	87	74	15	—2
30—40	128	73	77	12	—16	149	55	29	26	130	71	72	14	—15
40—50	104	53	57	5	—9	112	39	24	15	104	53	56	5	—8
50—60	84	31	37	12	—18	80	26	18	8	83	31	37	12	—18
Earth as a whole	128	77	68	9	0	132	46	27	19	129	68	56	12	0

The total radiation $(Q+q)$, which increases on the whole from high to low latitudes, reaches a peak not at the equator but rather at about latitude 20° N. and S.; this is clearly due to the greater cloudiness in the equatorial regions. Numerically a unit surface of land and sea receives practically the same amount of total radiation in a year for the earth as a whole.

In middle latitudes the radiation balance R increases rapidly with decreasing latitude. In the tropics ($\varphi < 30°$) the growth slows down and the values of R are fairly independent of latitude. On the average, in all zones, R is appreciably greater for sea surfaces than for land surfaces; this is a consequence of the smaller albedo and limited effective emission over oceans.

The loss of heat by evaporation (LE), as indicated earlier, is perceptibly greater over oceans than over land; maximum values of LE occur near the equator on land and in tropical latitudes (10–30°) on sea. Over the sea the values of LE drop slightly at the equator proper.

On the global average the turbulent heat transfer P_e from land surfaces is twice as large as that from ocean surfaces. Over oceans the values of P_e are somewhat greater in the high latitudes, while over land, on the contrary, the maximum values of P_e are recorded in tropical latitudes. It should be stressed here that, within the latitude interval under consideration (± 60°), the turbulent flux at the surface is everywhere positive and, therefore, both on land and on the oceans the flux of heat moves from the earth's surface towards the atmosphere (on the average).

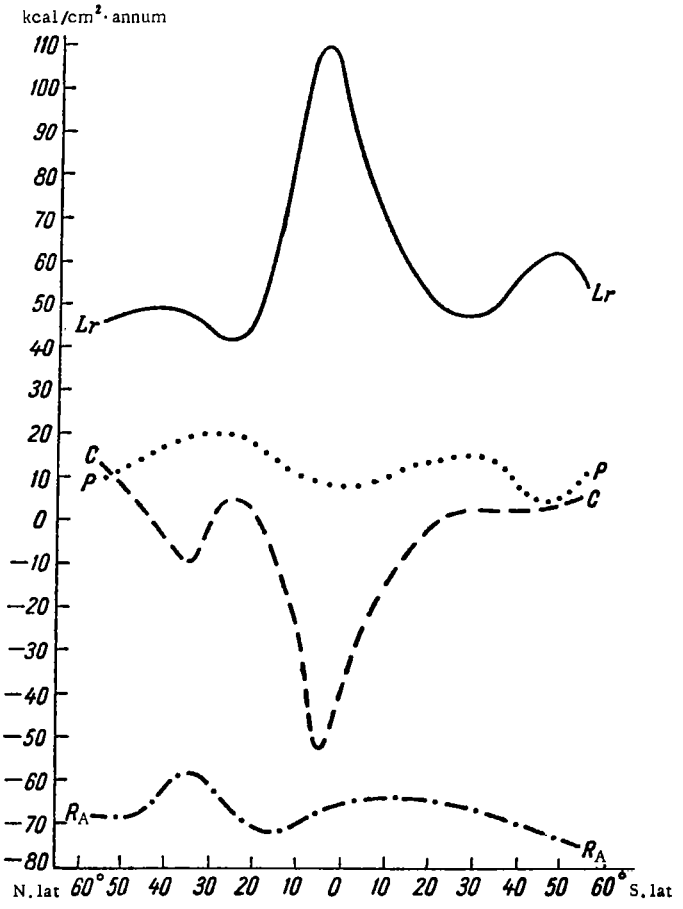


FIGURE 93. Thermal balance of the atmosphere*

Let us consider the term B_{ad}, which characterizes the flux between ocean surface and deeper layers due to ocean currents. The data in Table 55

* [The symbol C in this figure apparently stands for P_{ad}.]

show that, on the average, the oceans absorb heat in a zone extending roughly from latitude 20° S. to latitude 30° N. and give off this heat in higher latitudes.

The latitude distribution of the components of the atmospheric thermal balance on the annual average, given by equation (8'),

$$-R_A = P_e + Lr - P_{ad}$$

is shown in Figure 93.

From this figure one can see that, as already noted, the radiation balance of the atmosphere is negative at all latitudes. The influx of heat in the equatorial zone is determined to a very large degree by the heat released in the condensation of water vapor. In middle latitudes turbulent transfer also becomes important. Also apparent from the figure is the essential role of advection in the atmospheric balance; it effects the transfer of heat from equatorial and tropical regions to higher latitudes.

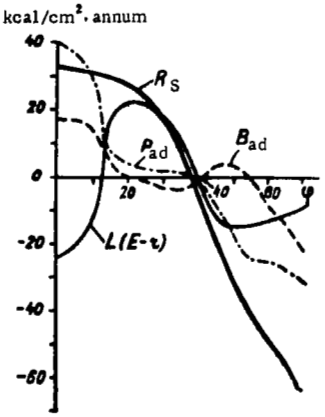


FIGURE 94. Thermal balance of the earth-atmosphere system

The latitude distribution of the components of the thermal balance of the earth-atmosphere system is of very great interest; it is given by the equation

$$R_\Sigma = P_{ad} + L(E-r) + B_{ad}$$

This distribution is shown in Figure 94. According to the figure, four latitudinal zones can be singled out in each hemisphere.

1. The equatorial zone, extending to latitude 10–15° north and south, characterized by a large positive radiation balance and considerable influx of heat from the moisture transport. All of this heat influx is expended in atmospheric and oceanic advection, for which the equatorial zone is the chief source of thermal energy.

2. A zone which includes the tropical regions and part of the subtropical regions, lying within 15–35° of latitude (approximately) in each hemisphere. This zone might be termed tropical. The distinguishing characteristic of the tropical zone is the large expenditure of heat in moisture transport. (this reaches considerable values when the radiation balance is positive), which decreases with increasing latitude. In these zones the fluxes of heat due to atmospheric and oceanic advection are insignificant.

3. Narrow transition zones at latitudes 35–40° in which the annual sums of the components of the thermal balance compensate each other.

4. Finally, zones with a negative radiation balance, lying at higher latitudes; here the absolute magnitude of the radiation balance decreases extremely rapidly with latitude. The influx of heat is due to atmospheric and oceanic advection, and also to moisture transport.

An important feature of the distribution of the components of the balance is that all the components play a substantial part in the transfer of heat between high and low latitudes and therefore none can be disregarded.

If one compares the latitude variation of oceanic and atmospheric advection one will see that the values of the latter are roughly twice those of the former. Thus the horizontal transport of heat in the atmosphere plays a greater role in its thermal balance than the horizontal transport of heat in the oceans. We note in addition that while the mean values of the radiation balance are only slightly different for the different oceans, they are substantially different for the individual continents.

§ 4. Thermal balance of the earth as a whole

We will now consider the mean thermal balance of the earth-atmosphere system for the earth as a whole. While a great deal of attention has been devoted to this question by many investigators, the calculations still retain a tentative flavor; moreover the schemes that have been proposed for the balance of the earth-atmosphere system differ from each other in certain details.

TABLE 53

Mean annual thermal balance of globe (after M.I. Budyko)

	Heat influx (kcal/cm^2· annum)		Heat loss (kcal/cm^2· annum)	
Earth's surface	Absorbed solar radiation	111 (44%)	Effective emission	43 (17%)
			Evaporation of moisture	56 (22%)
			Turbulent transport	12 (5%)
	Total	111 (44%)	Total	111 (44%)
Atmosphere	Absorption of short-wave radiation	39 (16%)	Balance of long-wave radiation	107 (43%)
	Condensation	56 (22%)		
	Turbulent transport	12 (5%)		
	Total	107 (43%)	Total	107 (43%)
Earth-atmosphere system	Received from sun	250 (100%)	Reflected	100 (40%)
			Emitted into outer space	150 (60%)
	Total	+250 (100%)	Total	−250 (100%)

We will present a calculation of the components of the thermal balance recently carried out by M. I Budyko, which is demonstrably the soundest. In this calculation it is assumed that the albedo of the earth as a whole is equal to 0.40 while the albedo of the earth's surface averages 0.14. Then out of the total flux of solar radiation, which can be taken to be 1000 kcal/cm^2· · annum, a unit surface at the top of the atmosphere receives on the average, owing to the spherical shape and the rotation of the earth, one quarter of the above amount, i.e., 250 kcal/cm^2· annum; the latter will be regarded as 100%. Of this 100 kcal/cm^2· annum (40%) is reflected back into outer space, and the earth's surface and atmosphere absorb the remaining 150 kcal/cm^2· annum (60%). Of the latter the earth's surface accounts for 111 kcal/cm^2· annum (44%) and the atmosphere for 39 kcal/cm^2· annum (16%).

Since the radiation balance of the earth's surface as a whole is equal to

68 kcal/cm^2· annum (27%; see Table 53), it follows that the effective emission at this surface is 111−68 = 43 kcal/cm^2· annum (17%).

The heat transfer due to evaporation and condensation amounts to 56 kcal/cm^2· annum (22%), and the turbulent heat transfer from the earth's surface to the atmosphere is 12 kcal/cm^2· annum (5%). As a result for the thermal balance we have the values cited in Table 53.

To permit detailed analysis the scheme under consideration (Table 53) has been broken down into three parts: thermal balance of the earth's surface, thermal balance of the atmosphere and thermal balance of the earth-atmosphere system.

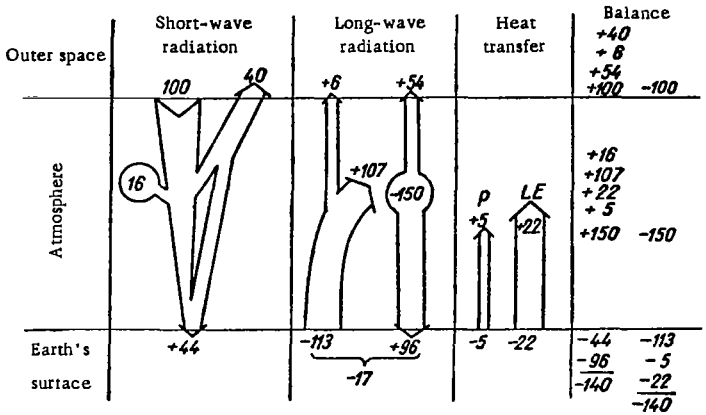


FIGURE 95. Thermal balance of the earth as a whole (components of the thermal balance are in percents)

The mutually compensatory fluxes of long-wave radiation have been left out of this table and also of the scheme given in Figure 95. For example, the long-wave emission of the atmosphere was not included in the influx of heat to the earth's surface and therefore it is the effective emission rather than the proper emission of the earth's surface that appears in the heat loss column. Also, no distinction is made in the table between direct and scattered solar radiation. As one can see from the table, the dominant role in the thermal balance of the earth's surface and atmosphere is played by the radiative transfer of heat and the heat transfer due to phase transformations of water. Turbulent transfer between the earth's surface and atmosphere is comparatively insignificant.

The thermal balance of the earth's surface was considered above; as to the balance of the atmosphere, it consists of the heat of absorbed short-wave radiation, which amounts to 39 kcal/cm^2· annum (16%), the influx of heat released in the condensation of water vapor and estimated at 56 kcal/cm^2· annum (22%), and the influx of heat due to turbulent transport, equal to 12 kcal/cm^2· annum (5%). In total the influx of heat is 39 + 56 + 12 = = 107 kcal/cm^2· annum (43%). The heat loss is due to effective emission, which also amounts to 107 kcal/cm^2· annum (43%). Thus the thermal balance of the earth as a whole amounts to ± 250 kcal/cm^2· annum, and of the 250 kcal/cm^2· annum delivered by the sun 100 kcal/cm^2· annum is reflected and 43 + 107 = 150 kcal/cm^2· annum is lost by long-wave emission. Somewhat different figures are cited for the northern hemisphere by Houghton and others, but the general character of these relations remains the same.

Part Four
MOISTURE REGIME OF THE ATMOSPHERE

The present moisture regime of the atmosphere is the product of a long interaction with the earth's surface.

The moisture regime of the atmosphere consists of and is governed by two processes: the evaporation of moisture from sea and land surfaces, which guarantees a continuous supply of water vapor to the atmosphere, and precipitation, by which moisture is returned to the earth's surface.

At the surface of the earth the total mass of water (not counting ground water) amounts to $1.3 \cdot 10^{+18}$ tons. This mass of water is almost entirely concentrated in the global oceans (99.7 %); polar ices account for 0.27 % and bodies of water in the interior (seas, lakes and rivers) for about 0.03 %.

The amount of moisture present in the atmosphere is roughly $1.3 \cdot 10^{13}$ tons, i.e., 0.001% of the global stores. Most of the moisture in the atmosphere is in the gaseous state (about 95 %); cloud particles (droplets of water and ice crystals) account for less than 5 % of the water contained in the atmosphere.

Most of the water vapor (> 90 %) is contained in the lower atmosphere (to heights of 8–10 km), where it constitutes 0.3–0.4 % of the mass of atmospheric air.

It is important to mention that the total amount of moisture received from the atmosphere per year in the form of precipitation, as also the evaporation, amounts to $5 \cdot 10^{14}$ tons for the entire globe; this is 40 times the total moisture content of the atmosphere.

This fact is a conclusive demonstration of the great intensity of the process of moisture exchange between the earth's surface and atmosphere.

Part Four deals with processes associated with the hydrologic cycle of the atmosphere.

Chapter 16

PHASE TRANSITIONS OF WATER

§ 1. Conditions of phase equilibrium

Water is the only substance encountered in the atmosphere in all three aggregate states: solid (ice), liquid (water) and gaseous (water vapor). This is due to the fact that the temperatures observed in the atmosphere are always lower than the critical temperature of water vapor, which is $T_{cr} = 647° K = 374° C$; the buoyancy at the critical point is $p_{cr} = 218$ atm.

Depending on partial pressure (buoyancy) and external conditions, the water vapor contained in the atmosphere may transform in part into the liquid or crystalline state.

Systems consisting of water in various aggregate states are often encountered in nature, for instance, in clouds. Moreover, the passage of water from one state to another — freezing and thawing, evaporation and condensation, etc.— is constantly observed in these systems. In such cases one speaks of different phases of water, taking phase to mean physically homogeneous parts of the system capable of transforming one into the other. For systems consisting of a single substance the concepts "phase" and "aggregate state" are identical.

In every system composed of several phases, molecular exchange leads to a constant transition of the substance from one phase to another (from one aggregate state to another). Often, however, a certain stable state is reached in the course of these processes in which all external manifestations of phase transitions cease and equilibrium between the phases is achieved in the system.

This equilibrium is essentially dynamic in character; it merely indicates that under these conditions a predominant transition of molecules from one phase to another does not take place — the mutual molecular exchange is balanced. A system which has reached phase equilibrium may remain in this state without any visible change as long as the external conditions remain constant.

The concept of the thermodynamic potential introduced by Gibbs is extensively employed to characterize phase equilibrium and conditions of phase transition. The thermodynamic potential is one of the most important characteristic functions that follow from the first and second laws of thermodynamics. We will give its definition. In order to do this, let us write down the first law of thermodynamics in the form

$$dQ = dU + Apdv, \tag{1}$$

where dQ is the amount of heat supplied to the system, dU is the internal energy and $Apdv$ is the work done against the external force of pressure.

Let us express dQ in terms of the entropy φ, making use of the relation (cf. Chapter 6) $d\varphi \geqslant \frac{dQ}{T}$, or $dQ \leqslant Td\varphi$; then

$$dU = Td\varphi + Apdv \leqslant 0, \tag{2}$$

where the equality sign applies to reversible processes and the inequality sign to irreversible processes; other symbols are as previously.

We will first consider reversible processes. If a process takes place isothermally ($T = \text{const}$) and at constant pressure ($p = \text{const}$), then expression (2) can be represented as the total differential of a certain function S

$$dS = d[U - T\varphi + Apv], \tag{3}$$

where the function

$$S = U - T\varphi + Apv \tag{4}$$

is called the thermodynamic potential.*

In all irreversible processes, as is proved in thermodynamics, there is a reduction in the thermodynamic potential ($dS < 0$); it remains unchanged ($dS = 0$) only when dynamic equilibrium prevails in a closed system and individual fluctuations are of a reversible nature.

If the thermodynamic potential decreases in all possible irreversible processes, it follows that it reaches a minimum in the case of phase equilibrium; this, indeed, is a necessary condition for phase equilibrium for specified values of the parameters T and p

For a system with one component the specific thermodynamic potentials of the phases should be equal for phase equilibrium. If the potentials are not equal the equilibrium between phases will be destroyed; and since in this case the system will tend to decrease its total thermodynamic potential, the phase with the higher potential will begin to pass over into the phase with the lower potential.

Let us make use of these relations in order to shed light on the conditions under which equilibrium is established between a liquid and its vapor. In this case, as was pointed out, the specific thermodynamic potential of the

* For greater detail on this see: Leontovich, Vvedenie v termodinamiku (Introduction to Thermodynamics), or Tverskoi, P.N. (editor). Kurs meteorologii (A Course in Meteorology).— Gidrometeoizdat. Leningrad. 1951.

two phases should be

$$S_{vap} = S_{liq} \quad \text{or} \quad dS_{vap} = dS_{liq}. \tag{5}$$

In the case of an isothermal process ($T = \text{const}$) we have, from (3), $dS = Avdp$. Consequently,

$$Av_{vap}\,dp = Av_{liq}\,dp,$$

or

$$A(v_{vap} - v_{liq})\,dp = 0. \tag{6}$$

Since $A = \text{constant}$ and $(v_{vap} - v_{liq}) \neq 0$, $dp = 0$ and $p = \text{const}$.

This means that isothermal equilibrium between a vapor and a liquid will occur only at a single well-defined pressure corresponding to the given temperature. The pressure required for equilibrium between the liquid and gaseous phases is simply the buoyancy of saturated vapor, which is usually denoted by E.

§ 2. Temperature dependence of buoyancy of saturated vapor over water and over ice

For equilibrium to be established in a water-vapor or ice-vapor system it is necessary that the buoyancy of vapor in the medium correspond to saturation. It has long been established experimentally that the buoyancy of saturated vapor increases sharply with rising temperature. In differential form this dependence, obtained from thermodynamic considerations, is given by the Clausius–Clapeyron equation derived in Chapter 6, § 10

$$\frac{dE}{E} = \frac{L}{AR_{vap}} \frac{dT}{T^2}, \tag{7}$$

where L is the heat of evaporation, E is the buoyancy of saturated vapor, A is the thermal equivalent of work, R_{vap} is the specific gas constant for water vapor and T is the temperature.

A more rigorous derivation of this equation can be obtained from the concept of the thermodynamic potential by using relations (5) and (3) for $U = \text{const}$. From (3) we have

$$dS = Avdp - \varphi dT. \tag{8}$$

Then the condition of equilibrium of the two phases

$$Av_{vap}dp - \varphi_{vap}dT = A_{vap}v_{liq}\,dp - \varphi_{liq}\,dT, \tag{9}$$

or

$$\frac{dp}{dT} = \frac{\varphi_{vap} - \varphi_{liq}}{A(v_{vap} - v_{liq})} = \frac{\Delta\varphi}{A\Delta v}. \tag{10}$$

The change in entropy in the given case is due to the loss of the heat of evaporation, so that

$$\Delta\varphi = \frac{\Delta Q}{T} = \frac{L}{T}.$$

In view of the fact that $_{liq} \ll v_{vap}$ and therefore $\Delta v \approx \frac{R_{vap}T}{F_{vap}}$, and that

$p_{vap} = E$ is the buoyancy of saturated vapor, we arrive at formula (7)

$$\frac{dp_{vap}}{dT} = \frac{dE}{dT} = \frac{LE}{AR_{vap}T^2}$$

To obtain the buoyancy of saturated vapor as a function of the temperature $E = f(T)$ it is necessary to integrate equation (7). We will take $T_0 = 273°$ and the corresponding value $E_0 = 6.1078$ mb as lower limits of integration. Then, assuming in the first approximation that $L = \text{const}$ we obtain

$$\ln \frac{E}{E_0} = -\frac{L}{AR_{vap}}\left(\frac{1}{T} - \frac{1}{T_0}\right). \tag{11}$$

But since

$$-\left(\frac{1}{T} - \frac{1}{T_0}\right) = \frac{T - T_0}{TT_0} = \frac{t}{T_0(273 + t)},$$

we have

$$\ln \frac{E}{E_0} = \frac{L}{AR_{vap}T_0}\frac{t}{273 + t}, \tag{12}$$

or

$$E = E_0 e^{\frac{L}{AR_{vap}T_0}\frac{t}{273+t}} = E_0 e^{C\frac{t}{273+t}}, \tag{13}$$

where

$$C = \frac{L}{AR_{vap}T_0}.$$

Passing over to common logarithms we have

$$E = E_0 10^{0.43\frac{Ct}{273+t}}. \tag{14}$$

If one assumes a linear dependence of the latent heat of evaporation on temperature then, as is known from experimental data and the theory of phase transformations, this dependence can be expressed as

$$L(T) = L_0 + (c_{p,vap} - c)(T - T_0) = L_0 - 0.57t,$$

where $L_0 \approx 597$ cal/g, c is the heat capacity of water (unity) and $c_{p,vap}$ is the heat capacity of water vapor, roughly equal to 0.43 cal/g.

Introducing the above in equation (7) we find

$$\int_{E_0}^{E} \frac{dE}{E} = \int_{T_0}^{T} \frac{[L_0 - a(T - T_0)]}{AR_{vap}T^2} dT, \tag{15}$$

whence, passing over to common logarithms, we have

$$\log E = \frac{K_1}{T} + K_2 \log T + K_3, \tag{16}$$

where K_1, K_2 and K_3 are constants; when their values are introduced the above expression becomes

$$\log E = 23.5518 - \frac{2937.4}{T} - 4.9283 \log T. \tag{17}$$

Let us use these formulas to calculate the buoyancy of vapor over water and over ice.

Buoyancy of water vapor over water. Since $A = 0.24 \cdot 10^{-7}$ cal/erg, and $R_{vap} = 4.6 \cdot 10^6$ erg/g·deg, assuming $L = L_0 \approx 597$ cal/g, we obtain the following numerical value for C

$$C = C_w = \frac{L_0}{AR_{vap}T_0} = 19.87.$$

Then formula (14) becomes

$$E = E_0 \cdot 10^{\frac{8.62t}{273+t}}. \tag{18}$$

Values of $E(T)$ calculated with the above formula do not coincide exactly with the experimental data. Magnus' empirical formula, put forward on the basis of experimental data, is

$$E = E_0 \cdot 10^{\frac{7.45t}{235+t}}, \tag{19}$$

where

$$E_0 = 6.1 \text{ mb } (4.6 \text{ mm}).$$

Results obtained from formulas (18) and (19) are compared with the experimental data in Table 54; the divergences between these values amount to hundredths of a millibar.

TABLE 54

Buoyancy of saturated vapor (mb) as a function of temperature

E	$t°$					
	-30	-20	-10	0	10	20
From (18)	0,53	1,27	2,87	6,11	12,32	23,70
From (19)	0,51	1,25	2,86	6,11	12,28	23,39
From experimental data	0,51	1,25	2,86	6,11	12,27	23,37

Buoyancy of vapor over ice. To determine the buoyancy of vapor over ice one should take instead of the heat of evaporation L_0 the heat of sublimation $L_i = L_0 + L_m$, where $L_m = 80$ cal/g is the heat of melting of ice. Then for the buoyancy of saturated vapor over ice surfaces E_i we obtain, by analogy with (11),

$$\ln \frac{E_i}{E_0} = -\frac{L_0 + L_m}{AR_{vap}} \left(\frac{1}{T} - \frac{1}{T_0} \right). \tag{20}$$

Performing the same transformations as above we find

$$E_i = E_0 e^{\frac{L_0 + L_m}{AR_{vap}T_0} \frac{t}{273+t}}, \tag{21}$$

or

$$E_i = E_0 10^{0.43 C_i \frac{t}{273+t}} = E_0 10^{0.43 \frac{L_0 + L_m}{AR_{vap}T_0} \frac{t}{273+t}}. \tag{21'}$$

The numerical value of C_i now becomes

$$C_i = \frac{L_0 + L_m}{AR_{vap}T} = 22.46. \tag{22}$$

Let us compare the buoyancy of saturated vapor over ice E_i and that over supercooled water E_w at the same temperature $t < 0$. From (13) and (21) we have

$$\frac{E_i}{E_w} = e^{\frac{L_m}{AR_{vap}T_0}\frac{t}{T}}, \tag{23}$$

where

$$C_i - C_w = \frac{L_m}{AR_{vap}T_0} = 2.66. \tag{24}$$

Since $t<0$ it is obvious from (23) that the buoyancy of saturated vapor over ice E_i is always less than the buoyancy of saturated vapor over supercooled water E_w at the same temperature. In view of (13) and (21) this difference can be expressed as

$$\Delta E = E_w - E_i = E_0\left[e^{C_w\frac{t}{T}} - e^{C_i\frac{t}{T}}\right].$$

From this, taking the derivative $\frac{\partial(\Delta E)}{\partial t}$, we find that ΔE has a maximum at

$$t = t_1\frac{T\ln\frac{C_w}{C_i}}{C_i - C_w} = \frac{T\log\frac{C_w}{C_i}}{0.43(C_i - C_w)}. \tag{25}$$

Introducing the numerical values $C_w = 19.80$ and $C_i = 22.46$ and taking T equal to 273° K, we find that $t \approx -12.6°$ C and the maximum value of ΔE amounts to about 0.269 mb.

Figure 96 shows this dependence of ΔE on temperature.

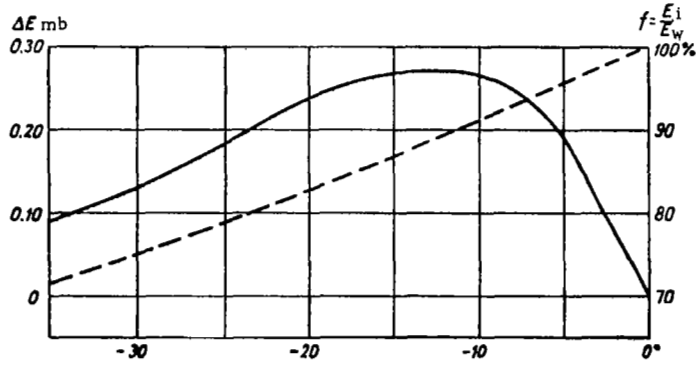


FIGURE 96. Difference ΔE between buoyancy of saturated vapor over water E_w and over ice E_i (solid curve) and relative humidity of air for saturation over ice (dashed curve)

The ratio $\frac{E_i}{E_w}$ is the relative humidity [for saturation over ice], which in meteorology is usually calculated with reference to the buoyancy of saturated vapor over a plane surface of distilled water. Since $E_i < E_w$, relative humidity [for ice-saturation] is always less than 100%.

For approximate calculation of the relative humidity for ice-saturation, an approximate relation can be obtained instead of expression (23). To

derive it we set $T = 266°K$; then, in view of (24), we write the following instead of (23)

$$\frac{E_i}{E_w} = e^{2.66\frac{t}{266}} = e^{0.01t}. \tag{26}$$

Expressing $e^{0.01t}$ as a series and retaining only terms of the second order of smallness, we obtain

$$\frac{E_i}{E_w} = 1 + 10^{-2}t + 10^{-4}\frac{t^2}{2} + \ldots, \tag{27}$$

or, expressing $\frac{E_i}{E_w}$ in percents,

$$\frac{E_i}{E_w} = 100 + t + \frac{t^2}{200} + \ldots \tag{28}$$

Table 55 gives the numerical values of $\frac{E_i}{E_w}$ as a function of temperature.

TABLE 55

Values of $\frac{E_i}{E_w}100\%$

$t°$	0	-5	-10	-15	-20	-25	-30	-40	-50
$\frac{E_i}{E_w}$(%)	100	95.2	90.8	86.4	82.3	78.3	74.6	67.8	61.9

As one can see from the data given in the table, the ratio of the saturation vapor pressure over ice to that over water decreases appreciably with falling temperature (i.e., the relative humidity [for ice saturation] decreases). Thus if at $t = 10°$, $E_i = 0.91E_w$, at $t = -40°$, $E_i = 0.68E_w$. One can also see from Table 55 that if the relative humidity is 100% over water, at negative temperatures (e.g., in supercooled clouds) it may frequently correspond to supersaturation over ice. Moreover, this supersaturation $\Delta f_i = 100 - f$ may reach several tens of percents.

In Figure 96 the function $f(T)$ is shown by the dashed line.

§ 3. Thermodynamic diagram of phase equilibrium

The temperature dependence of the buoyancy of saturated water vapor is represented graphically in Figure 97; the latter is the thermodynamic diagram of phase equilibrium of water.

The three states — water, ice and vapor — can occur simultaneously in equilibrium only for a single definite value of the temperature and pressure, namely at $t = 0.0076°$* and vapor pressure $e = p = 6.11$ mb. On the diagram the state of the system corresponding to these values of T and e is indicated by the letter O; this point is called the triple point.

The diagram shows several curves: OA, OC, OB and OB'.

Curve OA corresponds to the equilibrium between the liquid and vapor at positive temperatures (formula (15)). If at a certain temperature the buoyancy of water vapor is lower than the saturation pressure at the given

* *This temperature corresponds to the melting point of ice at the pressure* $p = 6.11$ mb; *at atmospheric pressure* $p = 1000$ mb the melting point of ice is 0°.

temperature, only the gaseous phase will be present (e. g., the state characterized by the point 1). If the vapor pressure is greater (e. g., the point 2) then equilibrium is possible only when part of the vapor condenses, leading to the state characterized by the point 3 (on curve OA). The line OA ends abruptly at $T = 374°$ K and $e = p = 218$ atm — the values of the critical temperature and critical pressure. Above lies the vapor region. Below this temperature the curve OA demarcates the boundary between the liquid and gaseous phases.

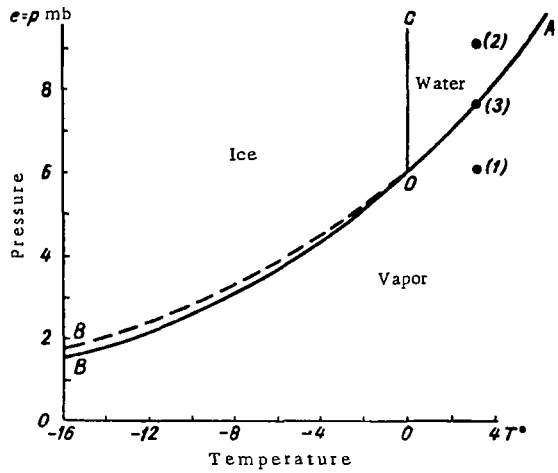


FIGURE 97. Phase equilibrium of water

At temperatures below 0° the evaporation line OA splits into two lines, OB and OB'. The former — (OB) — or sublimation line — corresponds to the buoyancy of saturation over ice (formula (18)), and the latter — OB' — is the line of saturation over supercooled water (formula (15)). The curve OB separates the regions of vapor and ice; it characterizes the state of equilibrium between ice and vapor. The vapor region lies below this line and the ice region above it. The dashed line OB' characterizes the same conditions for equilibrium of the vapor-supercooled water system.

The melting curve OC shows the relationship between the pressure and temperature of melting. For water, unlike most substances, it slopes to the left, i. e., the greater the pressure the lower the melting point.

Thus the curves drawn in Figure 97 demarcate three distinct regions: the vapor region below the lines OA and OB, the liquid water region between the lines OA and OC and the ice region (excepting supercooled water) between the lines OB' and OC. Equilibrium between two phases is possible only at values of e and T corresponding to the limiting curves.

If the state of a single-phase system is characterized by a point (with the values e and T) lying in the corresponding region of the diagram, the equilibrium is stable. If the system consists of two phases stable equilibrium between them is possible only for values of e and T corresponding to the boundary curves. If the point characterizing the state of the system (according to the parameters. e and T) lies outside these conditions of

equilibrium, e. g., the point 2 for vapor or the point 1 for water, then, from the graph, one may infer generally that the system is in an unstable state in this case and equilibrium does not prevail between the phases.

However, a whole series of states of the system corresponding to unstable equilibrium has been established experimentally. Thus in air cleaned of colloidal admixtures it is possible to obtain a supersaturated vapor without the formation inside it of a liquid phase. Vapor is said to be supersaturated when its buoyancy (pressure) is greater than that required for saturation at this temperature. Usually supersaturation, like relative humidity, is expressed in percents; since $\frac{e}{E} > 1$, $f > 100\%$. Sometimes the term supersaturation is applied also to the quantity $\Delta f = (f - 100)\% = \left(\frac{e}{E} - 1\right) \cdot 100\%$, i. e., the amount by which the relative humidity exceeds 100%.

Under certain conditions one can obtain water supercooled to temperatures of -30, $-40°$. A distinctive feature of unstable equilibrium is that the system may remain in this state for a long time and that certain additional conditions are necessary for the system to pass over into the stable state.

The diagram represented in Figure 97 characterizes phase equilibrium. As it permits graphical evaluation of the situation of a system, it is sometimes called the phase equilibrium graph. Using this graph one can easily determine the conditions required for stable equilibrium, but it is not possible to say anything about the additional conditions which govern the transition of the system from the unstable to the stable state at the same values of the temperature and pressure.

Chapter 17

EVAPORATION

Evaporation from oceans and land masses is the principal process by which water vapor enters the atmosphere. Evaporation uses up large quantities of heat, of the order of $3 \cdot 10^{23}$ cal/annum for the earth's surface as a whole, which amounts to about 25% of the energy delivered to the earth by the sun. The same amount of heat is released into the atmosphere in the condensation of water vapor. Thus, as we saw earlier, evaporation is the most important factor in the thermal balance of the atmosphere and earth's surface.

A great deal of space is devoted to the study of evaporation from water surfaces in hydrology and oceanography. Calculations of the rate of evaporation are tremendously important for the solution of many practical problems, particularly in the realm of reclamation, water supply, etc.

In the present chapter we will be concerned chiefly with the physical and meteorological aspects of this phenomenon and shall consider only the salient points of this highly complex problem.

§ 1. Process of evaporation from the kinetic theory standpoint

The following explanation of evaporation can be given on the basis of a kinetic theory: a certain number of molecules, having an energy greater than the mean kinetic energy at the given temperature and sufficient to overcome the attractive forces of the other molecules, are always breaking away from the surface of a liquid body. An amount of energy numerically equal to the heat of evaporation is used in doing this work.

As a result of this escape the space above the avaporating surface becomes enriched in vapor molecules. Due to diffusion some of the molecules of water vapor present in the thin layer of air directly above the evaporating surface are transported into the surrounding space; some are returned from this layer to the evaporating surface and are either reflected from this surface or recaptured by the liquid.

Schematically the process of evaporation can be represented as follows (Figure 98). From every square centimeter of evaporating surface $N_\uparrow$ molecules escape per second from the surface and $N_\downarrow$ molecules impinge upon the surface from the outside; of these N_{ref} are reflected and N_{abs} absorbed. The number of molecules actually lost by the liquid to the surrounding space is the evaporation N_{ev}. The latter is obviously given by

$$N_{ev} = N_\uparrow - (N_\downarrow - N_{ref}) = N_\uparrow - N_{abs} . \qquad (1)$$

Thus, the observed evaporation is given by the difference between the two streams of molecules escaping from and returning to the liquid. Numerically it is equal to the diffusive stream of water vapor.

Direct measurement of these streams is so complicated that this feat has yet to be carried out.

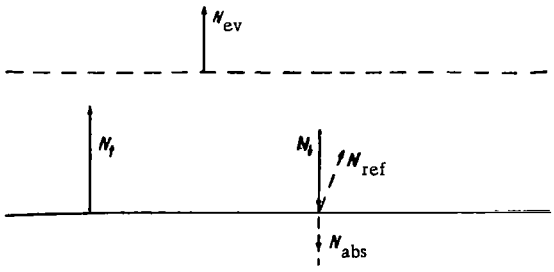


FIGURE 98. Evaporation (schematic)

A study of the molecular processes taking place near the evaporating surface enabled V. V. Shuleikin to construct a kinetic theory of evaporation. According to this theory the total number of molecules escaping across one square centimeter of surface per second is given by

$$N_1 = \frac{N}{\eta}\sqrt{\frac{kT}{2\pi m}}\,e^{-\frac{U}{kT}}, \qquad (2)$$

where N is the number of molecules per cubic centimeter of water, m the molecular mass, k Boltzmann's constant, T the temperature, U the work required to remove a molecule from the surface of the liquid, and η a certain correction factor which indicates what fraction of the molecules in the volume unit actually escape from the surface.

Since the numerical value of η is not given by Shuleikin his formula has not found practical applications so far.

However, certain of the molecular streams can be determined in another way. The stream of vapor molecules colliding with the surface of the liquid consists of molecules for which the component of velocity v [sic] is directed toward the evaporating surface. Their number can be found from Maxwell's law for the velocity distribution of gas molecules.

Let us assume that a unit volume contains n molecules (cm^{-3}); the fraction of molecules with velocities in the interval v_i and $v_i + dv$ [sic] is given by

$$dn = nf(v)\,dv,$$

where $f(v)$ is the Maxwell velocity distribution function $\left[f(v) = \frac{1}{n}\frac{dn}{dv}\right]$.*

In the time dt the evaporating surface will be reached by all molecules traversing a path $dh_i = v_i dt$ [sic]. The higher the volocity of the molecules the larger the layer from which they will arrive at the evaporating surface. The total number of molecules which will impinge upon the surface per

* Cf. a textbook of general physics.

square centimeter per second is

$$N_{\downarrow} = n\int_0^{\infty} f(v)\,v\,dv = n\left(\frac{m}{2\pi kT}\right)^{\frac{1}{2}}\int_0^{\infty} e^{-\frac{mv^2}{2kT}}\,v\,dv,$$

where m is the molecular mass, $k = \frac{R}{N_0}$ is Boltzmann's constant and T is the temperature.

Recalling that $d\left(\frac{mv^2}{2kT}\right) = \frac{m}{kT}\,v\,dv$ and taking the factor $\frac{kT}{m}$ outside the integral sign, we obtain

$$N_{\downarrow} = n\left(\frac{m}{2\pi kT}\right)^{\frac{1}{2}}\frac{kT}{m}\int_0^{\infty} e^{-\varepsilon}\,d\varepsilon = n\sqrt{\frac{kT}{2\pi m}}, \qquad (3)$$

where $\varepsilon = \frac{mv^2}{2kT}$.

This is the total number of molecules colliding with the surface; however, some of these may be reflected, and only a certain fraction, given by $\alpha = \frac{N_{abs}}{N_{\downarrow}}$, will enter the liquid. Thus the number of molecules returning to the liquid will be

$$N_{abs} = \alpha N_{\downarrow} = \alpha n\sqrt{\frac{kT}{2\pi m}}. \qquad (4)$$

Let us assume that in a thin layer directly above the evaporating surface $N_{\uparrow}$ is equal to N_{abs}; this is the condition of saturation for which the observed evaporation N_{ev}, from (1), is zero.

Evaporation is usually characterized not by the number of molecules that have escaped but by the mass of liquid evaporated. This quantity, referred to a unit surface per unit time, gives the rate of evaporation W, which is expressed in g $H_2O/cm^2 \cdot sec$. Obviously, this stream of water vapor amounts to

$$W = Nm. \qquad (5)$$

The expression for the rate of evaporation, by analogy with (1), can be written in the form

$$W_{ev} = W_{\uparrow} - W_{abs}. \qquad (6)$$

For the case of saturation, where $W_{ev} = 0$, obviously,

$$W_{\uparrow} = W_{abs} = \alpha W_{\downarrow}, \text{ where } W_{\downarrow} = mN_{\downarrow};$$

From (3), the number of molecules $N_{\downarrow}$ will be

$$N_{\downarrow} = n_{sat}\sqrt{\frac{kT}{2\pi m}}, \qquad (7)$$

where n_{sat} is the concentration of molecules in the layer above the liquid under conditions of saturation.

Bearing this expression in mind, we obtain

$$W_{\downarrow} = mN_{\downarrow} = mn_{sat}\sqrt{\frac{kT}{2\pi m}} = n_{sat}\sqrt{\frac{m}{2\pi kT}}\,kT. \qquad (8)$$

But $n_{sat}kT = E$ is the buoyancy of saturated vapor, and therefore

$$W_{\downarrow} = E\sqrt{\frac{m}{2\pi kT}}$$

and

$$W_{\uparrow} = W_{abs} = \alpha W_{\downarrow} = \alpha E \sqrt{\frac{m}{2\pi kT}}. \quad (9)$$

If in a layer of thickness h lying immediately above the evaporating surface the buoyancy of vapor $e_h = nkT$ is less than the saturation buoyancy ($e_h < E$), then, obviously, the stream of vapor absorbed from the outside by the surface will be given by

$$W'_{vap} = \alpha W'_{\downarrow} = \alpha e_h \sqrt{\frac{m}{2\pi kT}}. \quad (10)$$

Subtracting the above from (9) we obtain, in accordance with (6),

$$W_{ev} = W_{\uparrow} - W'_{vap} = \alpha (E - e_h) \sqrt{\frac{m}{2\pi kT}}. \quad (11)$$

This equation is called the Knudsen formula; it is frequently given in the following form

$$N_{ev} = \alpha (n_{sat} - n_h) \frac{\bar{v}}{4} \text{ mole/cm}^2 \cdot \text{sec} \quad (12)$$

where n_{sat} and n_h are the concentration of molecules for saturation and in the adjacent layer h, and $\bar{v}$ is the mean absolute velocity of the molecules for a Maxwell distribution

$$\bar{v} = \sqrt{\frac{8kT}{\pi m}}.$$

In view of the fact that from (3)

$$N_{\downarrow} = n \sqrt{\frac{kT}{2\pi m}},$$

one can easily obtain expression (12)

The value $\alpha = 0.036$ has been obtained from laboratory experiments. This means that out of 1000 molecules colliding with the surface of water only 36 are absorbed.

From equation (11), recalling (9) and (10), we find

$$e_h = \frac{W_{\uparrow} - W_{ev}}{\alpha \sqrt{\frac{m}{2\pi kT}}} = E \left(1 - \frac{W_{ev}}{W_{\uparrow}}\right).$$

The above shows that in the presence of evaporation ($W_{sat} > 0$) $e_h < E$, i. e., directly above the evaporating surface the buoyancy of vapor is always somewhat lower than the saturation pressure; moreover, the closer the rate of evaporation W_{ev} to evaporation in vacuum $W_{\uparrow}$, the less the buoyancy.

Evaluation of the magnitudes of these quantities shows that at (say) $T = 293°\,\text{K}$, $W_{\downarrow} = E \sqrt{\frac{m}{2\pi kT}} = 2 \cdot 10^{-2}$ g/cm$^2 \cdot$ sec and $W_{\uparrow} = W_{vap} = \alpha W_{\downarrow} = 7 \cdot 10^{-4}$ g/cm$^2 \cdot$ sec. At the same time the actually observed evaporation $W_{ev} = W_{\uparrow} - W_{\downarrow}$ in dry air at $t = 20°$ C under natural conditions amounts to $10^{-6} - 10^{-7}$ g/cm$^2 \cdot$ sec.

Thus the actual evaporation amounts to only a hundredth or thousandth part of the maximum possible evaporation in vacuum. In practice, therefore, it is always admissible to say that the buoyancy of vapor directly

above the surface of water $e_h \approx F$. At the same time from what we have said it is evident that the actual evaporation is limited not by molecular processes but rather by the rate of transport of water vapor in the atmosphere from the thin layer adjoining the surface. Thus it is necessary to consider the role of the process of diffusion. Let us denote the buoyancy of water vapor at a certain sufficiently large distance from the surface of the liquid by e; we shall assume that it remains constant in time. The buoyancy of vapor in the thin layer immediately adjoining the surface of the liquid, as we saw a little while earlier, is practically equal to the buoyancy of saturated vapor at the temperature of the surface.

In the case of completely immobile air the transport of molecules from the layer near the surface of the liquid into the surrounding space is effected by the process of molecular diffusion. This transport, which determines the observed rate of evaporation W_{ev}, can be expressed as the vertical flux of water vapor

$$W_{ev} = -D\frac{\partial a}{\partial z} = -D\rho\frac{\partial q}{\partial z}, \tag{13}$$

where D is the coefficient of molecular diffusion, and $a = \rho q$ is the volume concentration (density) of the water vapor (q being the specific humidity and ρ the density of air).

Recalling that $q = 0.622\,\frac{e}{p}$ and writing the last equation in finite differences, we obtain

$$W_{ev} = -0.622\,\frac{D\rho}{p}\,\frac{\Delta e}{\Delta z}. \tag{14}$$

If we denote $-0.622\,\frac{D\rho}{\Delta z}$ by A and assume that for $z \to 0$ $e \to E_{vap}$, where E_{vap} is the buoyancy of saturated vapor near the surface, we can write the above in the form

$$W_{ev} = \frac{A}{p}\,[E_{vap} - e] = A'\,[E_{vap} - e]. \tag{15}$$

For a more rigorous treatment of the process of diffusion of water vapor formula (15) would have to give the moisture deficit in terms of the molecule concentration (i.e., in terms of the density of the water vapor) rather than in terms of its buoyancy. However, for small temperature differences between the evaporating surface and the air equation (15) is sufficiently exact. It follows from this equation that: 1) the rate of evaporation is proportional to the moisture deficit $(E_{vap} - e)$ at the temperature of the evaporating surface, and 2) evaporation depends on the coefficient of diffusion of water vapor, the value of which, for immobile air at the temperature T and pressure p, is given by $D_{T,p} = D_0\,\frac{T}{T_0}\,\frac{p}{p_0}$, where $D_0 = (0.22 \pm 0.02)\,\mathrm{cm}^2 \cdot \mathrm{sec}$, $T_0 = 273°\,\mathrm{K}$ and $p_0 = 1000\,\mathrm{mb}$.

The dependence of the rate of evaporation on the difference $(E_{vap} - e)$ is very important. Evaporation occurs when the relative atmospheric humidity $f < 100\%$; the rate at which it takes place increases with rising temperature, due to a certain variation of the coefficient of diffusion D but chiefly because of the increase in the value of E_{vap} and consequent increase in the moisture deficit (for the same e). We note that the temperature of the evaporating surface may be lower than that of the surrounding air in

many cases; then the buoyancy E_{vap} of the vapor above this surface will be smaller than e, leading to condensation of the water vapor.

The dependence expressed in formula (15) for the rate of evaporation was obtained empirically on the basis of laboratory experiments at the beginning of the last century and is known as Dalton's law.

§ 2. Special features of evaporation in nature

The principal factors governing the rate of evaporation under natural conditions, according to Budyko, are the following:

1) the water supply, i. e., the availability of moisture for evaporation;

2) the influx of heat to be expended in evaporation, i. e., energy supply;

3) turbulent mixing, which leads to transport of the water vapor in the atmosphere.

In the presence of a water supply evaporation will depend on the character and state of the evaporating surface. For example, evaporation from the surface of oceans and seas is affected by the salinity of the latter, since the buoyancy of saturated vapor decreases over solutions. At low relative humidities this factor is not important but at a humidity of, say, 90%, the ratio of evaporation from sea water to that from fresh water amounts to 0.8, which decreases to only 0.6 at a humidity of 95%.

A reduction in the buoyancy of saturated vapor also occurs in evaporation from snow and ice surfaces. At low temperatures the buoyancy of saturation over ice is considerably lower than that over supercooled water, which in fact evaporates more rapidly than ice.

Conditions of evaporation are particularly complicated on land. Some areas are bare; some areas are covered by vegetation of assorted kinds; finally some areas are fairly wet while others are comparatively dry. The evaporation of water by plants, known as transpiration, is a complex physiologic process. It results in moisture being sucked out of the soil by the root system and subsequently evaporated from branches, leaves, flowers, and other parts, the surface of which can be very large. Accounting for the influence of vegetation is a very complicated affair.

In addition, when examining evaporation from land surfaces one must always consider whether the water evaporated from the surface will be continually replenished by moisture rising from underlying soil layers; this in turn depends on the humidity and structure of the soil, height of the water table, presence and character of the vegetation, and so forth. All this leads to broad variations in the observed rate of evaporation. Thus evaporation from dry soil is limited and may be close to zero, whereas from a thoroughly wetted surface evaporation will not be any less than over water; in the presence of vegetation — forest, shrubs, etc. — it may even be greater than this (e. g., in summer) due to evaporation from the very great exposed surface of plant parts.

The second factor on which evaporation depends, as we mentioned above, is the heat influx. Evaporation involves the expenditure of heat and if the evaporating surface does not receive heat its temperature will drop in the process. When this happens evaporation may not only decrease but, should the buoyancy of the saturated vapor E_{vap} over the surface reach the buoyancy e in the air, it may even stop. As a result the rate of evaporation

is largely dependent on the energy supply and is even sometimes determined entirely by the latter (i. e., by the amount of heat available for evaporation). Evaporation is considerable where heat and moisture are plentiful.

Finally, the third factor is the transport of water vapor in the atmosphere. It should be borne in mind that under natural conditions evaporation takes place not in calm air but rather in air which is in continual turbulent motion. Therefore the transport in the atmosphere of water vapor from an evaporating surface is determined chiefly by the intensity of turbulent mixing. This calls for replacing the coefficient of molecular diffusion D in formula (14) by the coefficient of turbulent diffusion.

However, turbulent transfer varies greatly with wind speed, increasing as the latter increases. Hence we have an obvious dependence of the rate of evaporation on wind speed. This dependence is highly complex. Indeed, intensive turbulent transfer leads to a reduction of the gradient of specific humidity and, from equation (13), the rate of evaporation is determined jointly by diffusion and the humidity gradient.

The influence of wind on the rate of evaporation has yet a further aspect: in the presence of wind there is a horizontal advance of vapor-laden air over the evaporating surface. This point becomes especially significant when large horizontal humidity contrasts form in the atmosphere and when evaporation is considered over small evaporating surfaces (e. g., reservoirs). In such cases more humid and less humid air masses advance alternately over the evaporating surface; when drier air advances, for instance, evaporation increases. This is particularly evident in the coastal areas. Obviously, in evaporation from a large, uniform tract the role of horizontal transport is reduced practically to nothing, since horizontal gradients will be close to zero and evaporation will take place almost entirely as a result of the vertical propagation of water vapor by turbulent diffusion.

It should be recalled that, in nature, all these factors, as well as other secondary factors not mentioned above, act simultaneously. Due to their diversity and complexity, separating and studying each factor individually, whether experimentally or theoretically, becomes a very difficult task.

§ 3. Evaporation and evaporation capacity

In meteorology it is customary to determine the mean rate of evaporation over a more or less prolonged interval of time: day, ten-day period, month, season or year; this rate, however, is expressed not in grams of moisture evaporated per square centimeter but in terms of the height of the evaporated layer of water, expressed in millimeters or centimeters. It is easy to compute that one kilogram of water spread over an area of one square meter will form a layer one millimeter thick; $1\ g/cm^2$ corresponds to a layer 10 mm thick.

Methods for measuring evaporation by means of so-called evaporators have long been in use; these are based on measuring the amount of water lost by evaporation from the surface of a quantity of water or soil sample artificially isolated in a special container. None of these methods, however, can give an exact value of the amount of water actually evaporated under natural conditions, since the conditions change drastically when the samples

are isolated artificially. Even if natural conditions of diffusion of vapor from the surface could be preserved completely, the rate of evaporation from the evaporator would obviously be modified by the disrutpion of conditions of heat transfer in the upper layers of water and soil. In the case of evaporation from the soil an additional drawback is the disruption of conditions of transport of moisture to the surface from deeper layers. This circumstance, noticed already in the earliest studies, has the effect that results depend on the size of the evaporator, method and place of installation, and a number of other factors. Numerous studies have been made in order to improve the design of the evaporators and remove the errors inherent in them.

In view of all these difficulties indirect methods in which the evaporation is not measured but calculated from observations of other elements with the help of the equation of the thermal or moisture balance have recently come into frequent use.

Since in nature evaporation takes place from a wide variety of surfaces (water, soil, plant, etc.), the rate of evaporation varies for the same meteorological conditions and cannot be determined uniquely. In view of this the concept of evaporation capacity is introduced in hydrology, meteorology and climatology. Originally this term designated evaporation from water surfaces, but now has a wider meaning.

Evaporation capacity has come to mean the rate of evaporation possible under the given conditions in the presence of a moisture excess. In other words, it characterizes the maximum possible evaporation allowed by the moisture and energy resources of the given locality. From this it follows that the actual evaporation is either equal to or less than the evaporation capacity. In deserts, for example, evaporation capacity is high but actual evaporation may be close to zero.

§ 4. Equation of turbulent diffusion of water vapor

The rate of evaporation depends strongly on the diffusion of water vapor. Let us therefore examine the general equation of transport of water vapor in a turbulent atmosphere. We will characterize the water vapor content of the atmosphere by the specific humidity $q = 0.622\ \frac{e}{p} = \frac{a}{\rho}$ = g H_2O/g of air, which is the ratio of the density of water vapor a to the density of air. Then for the flux of water vapor in the atmosphere in the vertical direction we can write

$$Q = -k_z \rho \frac{\partial q}{\partial z}, \tag{16}$$

where k_z is the coefficient of turbulent diffusion of water vapor, approximately equal to the coefficient of turbulence.

Reasoning in the same way as in the problem of turbulent transfer of heat (cf. Chapter 12, § 3), we obtain the equation of turbulent diffusion of water vapor:

$$\rho \frac{dq}{dt} = \frac{\partial}{\partial z}\left(k_z \rho \frac{\partial q}{\partial z}\right) + \frac{\partial}{\partial x}\left(k_x \rho \frac{\partial q}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_y \rho \frac{\partial q}{\partial y}\right). \tag{17}$$

Since $q = f(t, x, y, z)$, the total derivative of this function is

$$\frac{dq}{dt} = \frac{\partial q}{\partial t} + u\frac{\partial q}{\partial x} + v\frac{\partial q}{\partial y} + w\frac{\partial q}{\partial z},$$

where $u = \frac{\partial x}{\partial t}$, $v = \frac{\partial y}{\partial t}$ and $w = \frac{\partial z}{\partial t}$ are the components of the velocity along the coordinate axes.

We rewrite equation (17) as follows:

$$\frac{\partial q}{\partial t} = -\left(u\frac{\partial q}{\partial x} + v\frac{\partial q}{\partial y}\right) - w\frac{\partial q}{\partial z} + \frac{1}{\rho}\left[\frac{\partial}{\partial z}\left(k_z\rho\frac{\partial q}{\partial z}\right) + \right.$$
$$\left. + \frac{\partial}{\partial x}\left(k_0\rho\frac{\partial q}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_0\rho\frac{\partial q}{\partial y}\right)\right], \qquad (18)$$

taking $k_x = k_y = k_0$.

To allow for the outflow or inflow of water vapor to the volume of air due to phase transitions (evaporation and condensation) one must add the term $\frac{m}{\rho}$ to the right-hand side of the above; here m is the mass of water vapor (or water) condensed (or evaporated) per unit volume per unit time. We then obtain the equation of diffusion of water vapor in the atmosphere in its most general form

$$\frac{\partial q}{\partial t} = -\left(u\frac{\partial q}{\partial x} + v\frac{\partial q}{\partial y}\right) - w\frac{\partial q}{\partial z} + \frac{1}{\rho}\left[\frac{\partial}{\partial z}\left(k_z\rho\frac{\partial q}{\partial z}\right) + \right.$$
$$\left. + \frac{\partial}{\partial x}\left(k_0\rho\frac{\partial q}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_0\rho\frac{\partial q}{\partial y}\right)\right] \pm \frac{m}{\rho}. \qquad (19)$$

From the above it follows that at any point in the atmosphere the change in specific humidity results from:

1) transfer of vapor in a horizontal direction by the mean flow (advection of water vapor), expressed by the first term $\left(u\frac{\partial q}{\partial x} + v\frac{\partial q}{\partial y}\right)$;

2) transfer of vapor in the vertical direction by orderly vertical flows, given by the second term $-w\frac{\partial q}{\partial z}$;

3) turbulent mixing in the vertical and horizontal directions, given by the third term (in square brackets). We note that this process leads to equalization of the specific humidity;

4) evaporation of the water, or condensation of the vapor, contained in the volume under consideration (but not at the earth's surface), given by the term $\frac{m}{\rho}$.

The general solution of this inhomogeneous second-order partial differential equation presents very grave difficulties, the major one being that the coefficient of turbulent diffusion k_z depends on height and does not remain constant in time. In addition one needs to know the initial distribution (at $t = 0$) of specific humidity and initial wind profile (initial condition); also, the boundary conditions governing the variation of the specific humidity in time at the boundaries of the region under consideration must be specified. These initial and boundary conditions can be specified in different ways, corresponding to different kinds of concrete problems.

Works carried out along these lines by various authors (Jeffreys, Sverdrup, Sutton, Laikhtman, Timofeev, et al.) show that the problem can be solved in a more or less satisfactory manner only for an "infinite" or

"semi-infinite" surface of water, e.g., for an area in the middle of the ocean. However, the exploitation of the results obtained for practical calculations has proved to be a very complex and difficult matter.

Let us consider a few particular cases. To do this we simplify equation (19), taking the x-axis along the direction of the wind, so that $v = \frac{\partial y}{\partial t} = 0$. Then we set $w = 0$, and also assume that $k_z \gg k_0$, which gives us a basis for taking $k_0 = 0$. Then for the case where no phase transitions take place ($m = 0$), we write instead of (19)

$$\frac{\partial q}{\partial t} = \frac{1}{\rho} \frac{\partial}{\partial z}\left(k_z \rho \frac{\partial q}{\partial z}\right) - u \frac{\partial q}{\partial x}. \tag{20}$$

Integrating this equation over z we have

$$\int_0^z \frac{\partial q}{\partial t} dz + \int_0^z u \frac{\partial q}{\partial x} dz = k_z \frac{\partial q}{\partial z} + C, \tag{21}$$

where C is a constant of integration, proportional to the flow of water vapor near the surface, i.e., to the rate of evaporation. To determine it we note that for $z \to 0$ the two terms in the left-hand side of the above tend to zero. Consequently the constant C is given by

$$C = -k_z \frac{\partial q}{\partial z}\Big|_{z \to 0}.$$

But the quantity $-k_z \frac{\partial q}{\partial z}\Big|_{z \to 0} = -\frac{k_z}{\rho} \frac{\partial a}{\partial z}\Big|_{z=0} = \frac{W_{ev}}{\rho}$ gives the rate of evaporation and therefore

$$C = -k_z \frac{\partial q}{\partial z}\Big|_{z \to 0} = \frac{W_{ev}}{\rho}. \tag{22}$$

Now we write (21) in the form

$$W_{ev} = \rho\left[-k_z \frac{\partial q}{\partial z} + \int_0^z \left(\frac{\partial q}{\partial t} + u \frac{\partial q}{\partial x}\right) dz\right]. \tag{23}$$

In the right-hand side of the above the first term gives the magnitude of the stream of water vapor through the level z, the second term characterizes the influence of nonstationarity on the evaporation process, and the third term characterizes the influence of horizontal transport of water vapor. Evaluation of the numerical values of these terms shows that, in the presence of a uniform underlying surface, for the lowermost atmospheric layer ($z < 100$ m) the second and third terms are negligibly small compared with the first (roughly three orders smaller). This allows us to take the following approximate relation instead of (23)

$$W_{ev} = -\rho k_z \frac{\partial q}{\partial z}. \tag{24}$$

Calculations show that the error due to substitution of (24) for (23) does not exceed 10% for the air layer ranging over 10–100 m of altitude in daytime and 1–10 m at night.

Thus we arrive at the important conclusion that in the lowermost surface layer of the atmosphere the stream of water vapor, given by $-\rho k_z \frac{\partial q}{\partial z}$, can be regarded with sufficient accuracy as constant with

height. In view of this, integrating (24) over z, we find

$$W_{ev}\int_0^z \frac{dz}{k_z} = -\int_{q_{vap}}^{q_z} \rho dq = -\int_{a_{vap}}^{a_z} da = a_{vap} - a_z, \tag{25}$$

where q_{vap} and a_{vap} refer to the evaporating surface.

Hence we can write

$$W_{ev} = D'(a_{vap} - a_z) = D'\rho(q_{vap} - q_z), \tag{26}$$

where

$$D' = \frac{1}{\int_0^z \frac{dz}{k_z}}. \tag{27}$$

The coefficient D' is the integral characteristic of transfer and is called the coefficient of outward diffusion; it varies comparatively slightly with height. To calculate it one should adopt some hypothesis or other concerning the form of the function k_z for small z.

Taking the simplest linear dependence $k_z = k_1 z$, from (25) we obtain, for small altitudes,

$$W_{ev} = -k_1\rho\frac{q_2 - q_1}{\ln\frac{z_2}{z_1}} = -k_1\frac{a_2 - a_1}{\ln\frac{z_2}{z_1}}. \tag{28}$$

Comparing (26) and the above we find that

$$D' = \frac{k_1(q_1 - q_2)}{\ln\frac{z_2}{z_1}(q_{vap} - q_2)}. \tag{29}$$

Calculations show that on the average the values of D' amount to 1 — 1.5 cm/sec.

From data of numerous observations one can take $k_1 = 0.05u_1$, where u_1 is the wind speed at a height of one meter. Then from (28) we obtain approximately

$$W_{ev} = 0.05u_1\frac{a_1 - a_2}{\ln\frac{z_2}{z_1}}. \tag{30}$$

Expressions (30) and (28) make it possible to determine the rate of evaporation provided one knows the value of the coefficient of turbulence k and the values of the absolute or specific humidity at two heights: to measure the latter it is necessary to perform gradient observations.

The equation of the thermal balance can also be used to calculate evaporation. Indeed, we saw that in the equation of thermal balance of the surface, one of the components was the loss of heat in evaporation. Consequently, if all the other components are known one can determine the actual evaporation.

The equation of thermal balance was written in the form

$$R = P + B + LW, \tag{31}$$

where R is the radiation balance, P the heat exchange with the air, B heat transfer in the soil and LW the loss of heat in evaporation. As we noted earlier, all values are computed over a definite time interval t and referred to a unit surface.

Let us extract the evaporation from equation (31)

$$W=\frac{R-P-B}{L}. \tag{32}$$

This expression gives the general solution to our problem.

We will consider a particular solution due to M. I. Budyko and M.I. Yudin. We introduce into (31) the values

$$W=-k\rho\frac{\partial q}{\partial z} \text{ and } P=-k\rho c_p\frac{\partial\theta}{\partial z}, \tag{33}$$

where θ is the potential temperature.

Then

$$R-B=P+LW=-k\rho\left[c_p\frac{\partial\theta}{\partial z}+L\frac{\partial q}{\partial z}\right]. \tag{34}$$

But since from (33) $k\rho=-\frac{W}{\frac{\partial q}{\partial z}}$, we have

$$R-B=W\left[c_p\frac{\partial\theta}{\partial z}\frac{\partial z}{\partial q}+L\right],$$

or

$$W=\frac{R-B}{c_p\frac{\partial\theta}{\partial z}\frac{\partial z}{\partial q}+L}. \tag{35}$$

Passing over to finite differences (or integrating over z), we obtain

$$W=\frac{R-B}{L+c_p\frac{\theta_1-\theta_2}{q_1-q_2}}\approx\frac{R}{L+c_p\frac{\Delta T}{\Delta q}}. \tag{36}$$

It should be stressed that the above can be used only to determine the rate of evaporation at a definite instant, and mean (e. g., daily, monthly) values of the differences $\Delta T=\theta_1-\theta_2$ and $\Delta q=q_1-q_2$ must not be used for calculations with this expression.

Expression (36) is very convenient for calculating the evaporation since it does not contain the coefficient of turbulent exchange; one is only required to know the differences of temperature and specific humidity between certain altitudes.

Analysis of actual observational material shows that in most cases, especially over large time intervals, $R\gg B$ and $c_p\frac{\theta_1-\theta_2}{q_1-q_2}\ll L$; thus, for instance, for the mean annual evaporation the following relation holds

$$W=\frac{R}{L}. \tag{37}$$

This means that the mean annual evaporation in a given area is determined by the radiation balance for a comparatively large territory. This formula has proved to be sound for calculations of the evaporation from water surfaces and from soil surfaces in zones of abundant moisture.

§ 5. Empirical formulas for the calculation of evaporation

The determination of natural evaporation from both water and land surfaces is essential for the solution of a number of applied problems arising, for instance, in drainage, irrigation, and water supply. Many investigators

have therefore sought to obtain empirical relations linking evaporation to the major meteorological characteristics that govern it.

Formulas based on relation (15) have gained widest currency. These formulas express the dependence of the rate of evaporation on the moisture deficit and wind speed. They all have the form

$$W_{ev} = (E_{vap} - e) f(u), \tag{38}$$

where $(E_{vap} - e)$ is the moisture deficit; E_{vap} is reckoned for the temperature of the evaporating surface and $f(u)$ is a certain empirically established function which allows for the dependence of the rate of evaporation on wind speed (the so-called "wind factor").

A wide variety of expressions, more or less complex, has been suggested for the function $f(u)$. In most cases these have the form $f(u) = A + Bu$ (where A and B are numerical coefficients) or the form of a power $f(u) \sim u^n$, where the index n, according to various authors, lies between 0.5 and 1.

Since data on the temperature of the evaporating surface are often not available, formulas in which the moisture deficit is computed from the air temperature have been proposed. Formulas of this type become rather complicated when applied to evaporation from soil surfaces, and in certain of these a parameter which depends on the moisture is sometimes introduced.

The number of empirical formulas proposed by different authors is very considerable. Therefore we will not present them here, the more so as, like all empirical formulas, they are valid for the conditions under which the constants appearing inside them were determined but can produce substantial errors when used under other conditions. In most cases they can be used only to obtain tentative mean values.

§ 6. Results of observations of evaporation and evaporation capacity

The best available data on evaporation refer to ocean surfaces. One may assume on the average that 3–4 mm are evaporated per day from the ocean surface in the equatorial zone, 1–2 mm in middle latitudes and still less in polar regions. Evaporation from the surface of inland seas and basins varies strongly with local conditions and time of year. In the Aral Sea, for instance, where the daily average is about 3 mm, evaporation reaches 10 mm per day in summer and decreases to 0.1 mm in winter. Evaporation from small reservoirs, lakes, ponds, pools and so forth is also highly variable.

Evaporation from the soil surface, as we saw earlier, depends very strongly in each area on the humidity of the soil, presence of vegetation and many other factors. In many areas, therefore, the evaporation may be very different from the evaporation capacity. As an example one might mention that in the Sahara, where evaporation from an open container with water could amount to 4000 mm per year, the actual evaporation is entirely negligible; similar conditions obtain in the Tashkent area, where the evaporation capacity can be estimated at about 2000 mm per annum. Equally strong is the influence of vegetation, the presence of which reduces direct

evaporation from the soil to some extent (as a rule). When evaporation from plant surfaces is also considered, this so-called total evaporation may prove to be considerably larger than that from a bare soil surface.

As to the actual evaporation, it is determined by both the water and the energy supply and, according to calculations, amounts to about 150 mm/annum in the northern zone of the European USSR (tundra). It increases southward, reaching 300–400 mm/annum in the central zone; in the south-eastern zone of the European USSR and in Central Asia it decreases again to values below 200 mm/annum. We note that in the tundra the low evaporation is due to a lack of heat and in the southern regions to a lack of moisture.

Results of calculations of the global distribution of evaporation indicate that evaporation decreases polewards from the equatorial regions. In the northern hemisphere, however, the maximum evaporation per year is observed in the tropics and not at the equator. Differences are perceptible in the evaporation from land and sea surfaces. While in the equatorial zone evaporation is about the same from land and sea, in the tropics and middle latitudes evaporation from the sea is greater than from land surfaces. This can be seen from Table 56, which gives the rates of evaporation (after M. I. Budyko) for various latitude zones of the northern hemisphere.

TABLE 56

Zonal distribution of evaporation (cm/annum)

	Latitude (°)							
	0–10	10–20	20–30	30–40	40–50	50–60	60–90	0–90
Land	112	57	37	41	37	23	10	41
Oceans	110	135	130	115	70	60	15	101
Northern hemisphere	110	114	95	83	53	39	12	77

The latitude distribution of evaporation is roughly the same in the southern hemisphere but due to the fact that the area occupied by the oceans is greater there the mean evaporation for the entire hemisphere amounts to a layer of about 125 cm/annum. Globally the average evaporation is about 100 cm/annum.

The daily march of evaporation is mostly parallel to the march of temperature. In daytime, when the temperature rises and turbulent transfer intensifies, the rate of evaporation increases. At night as the relative humidity drops and turbulent transfer decreases the evaporation decreases; it changes only slightly during the night and is close to zero.

In the summer months the daily march of evaporation is more clear-cut than in winter. In the annual march the maxima of both evaporation from land surfaces and evaporation capacity are recorded in summer (June, July, sometimes May) and the minima in winter (December–January). For permanently ice-free seas the maximum of the annual march occurs in the autumn-winter period.

§ 7. Distribution of water vapor in the atmosphere. Daily and annual march of atmospheric humidity

Evaporation from the earth's surface is practically the only process which supplies water vapor to the atmosphere. In the present section we will briefly consider the distribution of water vapor in the atmosphere, which is determined jointly by vapor transport and condensation processes.

TABLE 57

Mean values of e and f at various latitudes

	φ (° N.)						
	5	15	25	35	45	55	65
t (°)	25.5	25.4	21.9	15.3	8.7	1.2	-7.0
e (mb)	25.3	22.9	18.4	12.9	9.3	6.5	4.1
f (%)	79	75	71	70	74	78	82

The global distribution of water vapor can be pictured by generalizing results of measurements of atmospheric humidity. One may note that, characterizing the humidity of air by its buoyancy, on the whole the buoyancy of water vapor e resembles the temperature in its distribution: the maxima are recorded in the equatorial regions, and there is a poleward decrease. From Table 57, which gives the mean buoyancy of water vapor e and relative humidity f at various latitudes, it is seen that the former

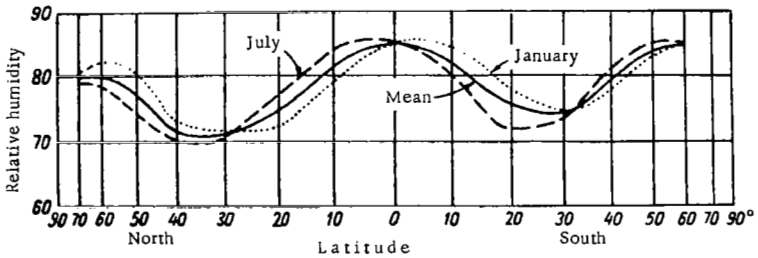


FIGURE 99. Zonal variation of relative humidity

reaches 25 mb in the equatorial zone and decreases to 4–5 mb at a latitude of 65–70° N. Obviously, the poleward drop in the humidity is smaller in summer and greater in winter. In winter at latitudes of 65–70°, under low-temperature conditions (–20° and lower), the buoyancy of vapor amounts to only 1 mb.

The latitude variation of the relative humidity f is more complicated. Typical of the temperate zone is a rise in relative humidity toward higher latitudes, especially in winter. This is clearly illustrated by Figure 99.

Atmospheric humidity also varies to some extent along parallels of latitude. The highest absolute and relative humidities are naturally recorded over oceans; as one moves inland there is a general decrease in

humidity. But even over continents local conditions create great nonuniformities in humidity distribution.

The question of the vertical distribution of atmospheric humidity is difficult to solve theoretically due to the complexity of the conditions which govern this distribution. The empirical formula

$$e_z = e_0 \cdot 10^{-\beta z},$$

quoted in Chapter 2 characterizes the mean variation of the buoyancy of water vapor to a height of 3–4 km. On the basis of numerous aerological observations this formula was corrected and extended to greater altitudes; a function of the following form was obtained:

$$e_z = e_0 \cdot 10^{-az - bz^2}, \tag{39}$$

here a and b are empirical coefficients.

Inserting the numerical values of the coefficients, the above formula may be written as follows for Western Europe:

$$e_z = e_0 \cdot 10^{-\frac{z}{8} - \frac{z^2}{48}}, \tag{40}$$

where z is expressed in kilometers.

On the basis of observations in the Moscow area A. Kh. Khrgian proposed the following formula for the vertical decrease of the specific humidity:

$$q_z = q_0 \cdot 10^{-\frac{z}{11.8} - \frac{z^2}{62.1}}. \tag{41}$$

where z is again in kilometers.

Comparing formulas (40) and (41) one can see that the decrease in the specific humidity q is slower than the decrease in the buoyancy of water vapor e. This is understandable, since $q = 0.622\frac{e}{p}$, and e decreases more rapidly than p.

The absolute humidity, i. e., volumetric concentration of water vapor a, decreases with altitude in a similar manner. It should be emphasized that everything we have said is valid only on an average. In each actual case there may be considerable deviations and the vertical variation of humidity with altitude, despite the general decrease, may be irregular.

In general, the relative humidity also decreases with height, but this decrease is considerably less regular.

In the surface layer of the atmosphere the variation of the vapor pressure is similar to the temperature variation; the march of the relative humidity, on the other hand, is the opposite of the temperature march. Moreover, the periodic changes in the humidity of this layer are affected directly by the vertical transport of water vapor. When transport in the surface layer is feeble the water vapor content near the earth's surface will be higher and its variation in time will depend only on evaporation from the underlying surface. Intensive transfer reduces the absolute humidity in lower layers and leads to enrichment of the higher layers. All this explains the peculiarities of the daily march of humidity.

Two types of daily march of the vapor pressure and specific humidity are distinguished (Figure 100).

The first type is simple and entirely analogous to the march of temperature: the diurnal maximum occurs in daytime at the same time as the temperature maximum, and the minimum before sunrise; in general the

amplitude increases together with the temperature amplitude. This type is observed when vertical transfer is insignificant and evaporation intense. It is recorded at places where an abundance of moisture ensures the possibility of continuous evaporation. This kind of daily march is usually present over extensive water surfaces (seas) and over continents in winter.

The second type of daily march of e or q has the form of a double wave (Figure 100) and is characterized by two maxima, at about 9-10 and 20-21 hrs, and two minima, in the early morning and during the period of greatest turbulence in the afternoon hours. This type of daily march is ordinarily observed on continents in summer.

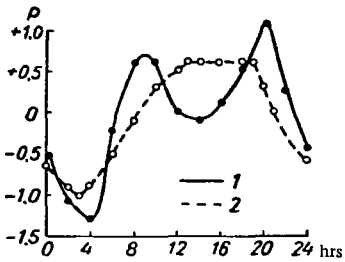


FIGURE 100. Daily march of vapor pressure

1-on clear summer days; 2-on all autumn days.

As one moves upwards the double-peaked curve of the daily march gradually changes into the simple one, and at a height of 2-3 km the daily oscillations vanish. The same process of vertical transfer is responsible for an increase in humidity higher up in daytime. The height at which the daily march changes character depends on the intensity of turbulence; on the average it is about 0.5 km at latitudes of 50-60°.

The relative humidity of the lower air always displays a daily march roughly opposite to the temperature march. Under continental conditions the daytime rise of the relative humidity is particularly marked in summer. In this case both the reduction in e due to transfer and the increase in E due to the temperature rise act in the same direction — to reduce f. Therefore the amplitude of daily oscillations of f is large on continents and far smaller over water surfaces.

In the daily march of relative humidity the maximum occurs before sunrise and the minimum at about 15-16 hrs.

In conformity with the reversal of the daily march of absolute humidity, a reversal of the march of relative humidity occurs as one rises over the surface. The daytime rise in the absolute humidity higher up results in a rise in the relative humidity as well.

The annual march of both absolute and relative humidity is simple: the absolute humidity repeats the pattern of the temperature, and the relative humidity is opposite to it. The values of the annual amplitudes also correspond, on the whole, to the values of the temperature amplitudes.

In summer the absolute humidity is maximum and the relative humidity minimum, and in winter vice versa. Thus in northern latitudes the relative humidity is maximum in winter and amounts at this time to 80-90%; in the summer months it decreases to 60-70%. In the dry south-eastern districts of the European USSR it drops to 40-50%. The buoyancy of water vapor, on the contrary, is low in winter (2-3 mb) and considerably higher in summer; mean values in the summer months reach 12-15 mb, so that annual amplitudes amount to 10-12 mb.

The annual march of humidity depends to a great extent on physico-geographic conditions.

The annual humidity oscillations decrease with altitude and become insignificant in the upper troposphere.

Chapter 18

FORMATION OF THE LIQUID AND SOLID PHASE OF WATER IN THE ATMOSPHERE

§ 1. Condensation and sublimation of water vapor

In addition to the process of evaporation of water and ice, the converse process — transition of water vapor into the liquid (condensation) or directly into the solid (sublimation) state — is always taking place in nature.

Condensation and sublimation of water vapor take place both on the earth's surface and objects lying at the surface, and in the atmosphere. The presence of a liquid or solid phase of water favors the initiation of these processes. For condensation (sublimation) of water vapor to take place it is then sufficient that the concentration (buoyancy) of water vapor in the air be greater than over the surface of the water or ice (or snow).

The situation becomes far more complicated in the case where liquid or solid particles of water are not present in an atmosphere containing water vapor. In this case 'embryos' of the new phase will first appear and the newly formed embryos must have a tendency to grow, rather than be destroyed, in the given system.

The formation of "viable" embryos of a new phase can occur as a result of density fluctuations of the vapor in an initially homogeneous system. Complexes of molecules which become stable and act as embryos of water droplets under definite conditions, may be formed in the system during such fluctuations.

It follows from the theory of spontaneous condensation that the probability for the formation of embryos is proportional to the quantity $e^{-\frac{W_{em}}{kT}}$, where $W_{em} = \frac{4}{3}\pi r_{em}^2 \sigma$ is the work required for the formation of an embryo, r_{em} is the radius of a viable embryo, σ is the surface tension, k is Boltzmann's constant and T is the temperature.

The number of such embryos I formed per cm^3/sec is given by the relation

$$\log I = \log C - \frac{W_{em}}{2.3\,kT}, \qquad (1)$$

where C is a certain coefficient proportional to the number of molecules and dependent on temperature and supersaturation $\left(\frac{e}{E}\right)$.

To derive the expression $W_{em} = \frac{4}{3}\pi r_{em}^2 \sigma$, let us consider the change in the thermodynamic potential upon formation of water droplets in a supersaturated vapor. We will assume that one droplet of liquid of radius r having the surface energy $F = 4\pi r^2 \sigma$ (the energy σ refers to a unit surface) was formed in the gaseous phase. For water $\sigma = 75$ dyn/cm = 75 erg/cm^2.

The thermodynamic potential of the system in the presence of a single droplet is

$$S = S_{vap} m_{vap} + S_{liq} m_{liq} + F,$$

and in the absence of the droplet

$$S_0 = S_{vap} m = S_{vap} (m_{vap} + m_{liq}),$$

where m is the total mass of the system.

Hence

$$\Delta S = -(S_{vap} - S_{liq}) m_{liq} + 4\pi r^2 \sigma.$$

Since

$$S_{vap} - S_{liq} = \frac{2\sigma}{\rho_w r} \text{ and } m_{liq} = \frac{4}{3}\pi\rho_w r^3,$$

we have

$$\Delta S = -\frac{2\sigma}{\rho_w r}\frac{4}{3}\pi\rho_w r^3 + 4\pi r^2\sigma = 4\pi\sigma\left(-\frac{2}{3}r^2 + r^2\right) = \frac{4}{3}\pi r^2\sigma, \quad (2)$$

which corresponds to the work done in the formation of the embryo.

Under natural conditions the formation of embryonic water droplets and ice crystals in a supersaturated vapor takes place on condensation nuclei ($r \approx 10^{-5} - 10^{-6}$ cm), which are always present in sufficient amounts in the atmosphere.

In the study of vapor condensation an important role was played by experiments in a Wilson cloud chamber, in which the process of condensation of water vapor can be reproduced under laboratory conditions similar to atmospheric conditions. These widely known experiments show that a cloud will form in the chamber only if a certain so-called critical supersaturation is created inside it.

In moist air carefully purified of suspended foreign particles a small number of comparatively large drops will be formed only for an adiabatic expansion corresponding to eight-fold supersaturation, i.e., for $\frac{e}{E}100 = 800\%$. At lower supersaturations the formation of cloud is not observed. When gas ions are present in the air cloud is formed at four-fold supersaturation ($(\frac{e}{E})100 = 400\%$) on negative ions and at six-fold supersaturation ($(\frac{e}{E})100 = 600\%$) on positive ions. In experiments with ordinary (nonpurified) air cloud is formed at very low supersaturations; in addition, a large number of minute droplets appears.

All these facts, established already at the end of the preceding century, subsequently received a theoretical explanation. Thus Krastanow, and later Frenkel and others, showed that the work done in forming the embryo of the new phase W decreased in the presence of nuclei by the quantity W_{nuc}, which depends on the size of the nucleus. That is, in calculations one should take not W_{em} but $(W_{em} - W_{nuc})$. Therefore, especially for $W_{em} \approx W_{nuc}$, the presence of condensation nuclei eases the conditions of droplet formation considerably, and low supersaturations become admissible.

It has been established experimentally that the formation of the solid phase of water in the atmosphere takes place in two ways: firstly, by direct sublimation (bypassing the liquid phase) on nuclei, or, at very low temperatures, spontaneously; secondly, by freezing of supercooled droplets. This question will be considered in greater detail in § 6.

§ 2. Vapor pressure over the surface of droplets

When considering the question of vapor pressure over the surface of droplets one must account for the influence of a number of factors, the most important of which are: 1) the curvature of the surface, 2) the presence of electric charge on the droplet, and 3) the presence in the droplet of dissolved hygroscopic impurities.

Thus, the buoyancy of vapor over droplets is a function of several quantities $E = f(r, q, k, T)$.

As initial value we will take the buoyancy of saturated vapor over a plane surface ($r = \infty$) of distilled water. We denote this quantity by E_∞. Then the buoyancy E of vapor over a droplet at a certain unchanged temperature can be written in the form

$$E = E_\infty + \Delta E_r - \Delta E_q - \Delta E_s, \tag{3}$$

where ΔE_r takes into account the influence of the curvature of the surface, ΔE_q that of the electric charge and ΔE_s that of the concentration of the solution.

Let us consider the dependence of E on these factors.

1. Influence of curvature of surface. It is well known that the buoyancy of vapor over a convex surface is greater, and that over a concave surface smaller, than over a plane surface of water. This is due to the change in the surface energy of the system.

Denoting by E_r the buoyancy of vapor over a surface of radius of curvature r, one can write

$$E_r = E_\infty \pm \Delta E_r. \tag{4}$$

From physics it is known that the relation between E_r and E_∞ is given by Kelvin's formula

$$\ln \frac{E_r}{E_\infty} = \frac{2\sigma}{\rho_w R_{vap} T} \frac{1}{r} = c_r \frac{1}{r}, \tag{5}$$

where σ is the surface tension at the water-vapor interface, ρ_w is the density of water (droplet) and R_{vap} is the gas constant of water vapor.

The surface tension depends on the temperature in a fairly complicated manner. The values of σ at different temperatures, and also the values of $c_r = \frac{2\sigma}{\rho_w R_{vap} T}$, are given in Table 58.

TABLE 58

Values of σ (dyn/cm) and c_r (cm)

t (°)	−5	0	10	20	30
σ (dyn/cm)	76.4	75.6	74.2	72.8	71.2
$c_r \cdot 10^7$ (cm)	1.1	1.2	1.2	1.3	1.3

We express (5) in the form

$$E_r = E_\infty e^{\frac{c_r}{r}}. \tag{5'}$$

Since on the average c_r is of the order of 10^{-7} cm, for droplets with

$r > 10^{-7}$ cm it is more convenient to use an approximate formula, which can be obtained from the above by expanding $e^{\frac{c_r}{r}}$ in series and retaining the first terms of the expansion

$$E_r = E_\infty \left[1 + \frac{c_r}{r} + \ldots\right] = E_\infty \left[1 + 1{,}2 \cdot 10^{-7} \frac{1}{r} + \ldots\right]. \quad (6)$$

Consequently

$$\Delta E_r = E_\infty \frac{c_r}{r}. \quad (7)$$

Table 59 gives the values of $\frac{E_r}{E_\infty}$ in percents, calculated at $t = 0°$ for several droplet sizes.

TABLE 59

r (cm)	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}
$\frac{E_r}{E_\infty}$ (%), from (5')	295	112.8	101.2	100.1	100.01
$\frac{E_r}{E_\infty}$ (%), from (6)	220	112.0	101.2	100.1	100.01

As one can see from the data, over small droplets ($r < 10^{-5}$ cm) very large supersaturations are necessary to forestall evaporation: in practice the influence of curvature becomes insignificant only for $r > 10^{-4}$ $\left(\frac{E_r}{E_\infty} < 0.1\%\right)$.

In deriving formula (6) it was assumed that the surface tension σ was independent of the droplet radius: this is true only for droplets of radius $r > 10^{-6}$ cm. Consequently, formula (6) can be regarded as relatively exact for $r \geq 10^{-6}$ cm.

2. Influence of electric charge. For charged droplets the equilibrium buoyancy of vapor diminishes, since in this case the influence of electric forces comes into play.

If the charge of the droplet is equal to n elementary charges e, then the buoyancy of saturated vapor E over such a droplet, according to J. J. Thomson, is related to E_∞ by

$$\ln \frac{E_{r,q}}{E_\infty} = \frac{1}{\rho_w R_{vap} T}\left[\frac{2\sigma}{r} - \frac{n^2 e^2}{8\pi r^4}\right]. \quad (8)$$

From this it follows that

$$\Delta E_{r,q} = E_\infty \frac{c_r}{r} - E_\infty \frac{c_q}{r^4} \text{ and } \Delta E_q = E_\infty \frac{c_q}{r^4}, \quad (9)$$

where $c_q = \frac{n^2 e^2}{8\pi \rho_w R_{vap}}$ which, for $n = 1$, is $7.5 \cdot 10^{30}$ cm^4.

Insertion of a correction for charge is meaningful only for droplets with $r \leq 10^{-7}$ cm.

Making use of the calculated values of c_r and c_q, instead of (9) one can write

$$E_{r,q} \approx E_\infty \left[1 + \frac{1.2 \cdot 10^{-7}}{r} - \frac{7.5 \cdot 10^{-30}}{r^4}\right].$$

The corrections for the curvature of the surface and influence of electric charge are of opposite signs. For very small values of r the correction

for the influence of electric charge may be greater in magnitude than the correction for the curvature of the surface. Then the buoyancy of vapor required for equilibrium over such a droplet will be less than E_∞. But for larger droplets ($r > 10^{-7}$ cm) the first term is always more significant and a buoyancy greater than E_∞ is required for equilibrium.

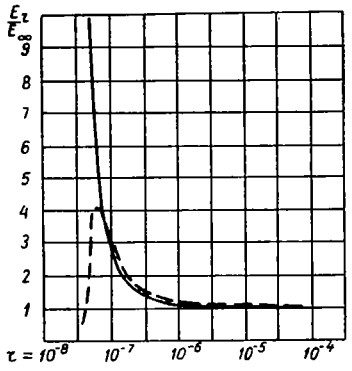


FIGURE 101. Buoyancy of vapor over droplets as a function of droplet radius

The dependence of $\frac{E_{r,q}}{E_\infty}$ on the droplet radius r is represented graphically in Figure 101; log r and $\frac{E_r}{E_\infty}$ are plotted on the coordinate axes. The solid curve represents this function without allowing for the charge and the dashed curve allowing for one elementary charge. As one can see, in the presence of charge it is necessary for the supersaturation in the air to be four-fold, i.e., the relative humidity must amount to 400% (peak of curve); then the particles will be viable and will be able to grow further at lower supersaturations. For $r > 10^{-7}$ cm the influence of electric charges practically vanishes.

3. Influence of hygroscopic impurities (salts and acids) dissolved in the droplet. Being a good solvent, water in nature always contains impurities. Thus sea water contains up to 40 g of dissolved salts per liter, well and spring water up to one gram, and rain water and snow usually 7–10 mg of salt per liter of water.

Admixtures of salts and acids reduce the buoyancy of saturated vapor over water. In the presence of dissolved impurities the buoyancy of saturated vapor over a plane surface of the solution E_s is

$$E_s = E_\infty - \Delta E_s, \tag{10}$$

where ΔE_s is a correction which depends on the concentration of the solution K.

For nonvolatile substances one can assume that soluble impurities are contained only in the liquid phase and are absent in the gaseous phase; but then the concentration of the solution increases when the molecules of the solvent transform into vapor and decreases upon condensation.

The concentration of the solution K is usually expressed as the ratio of the number of moles of dissolved substance n to the number of moles of the solution $N+n$, i.e.,

$$K = \frac{n}{N+n}, \tag{11}$$

where $n = \frac{m}{\mu_1}$, $N = \frac{M}{\mu_2}$ (m and M being the masses of the dissolved substance and solvent, and μ_1 and μ_2 their molecular weights).

According to Raoult's empirical law

$$E_s = \frac{N}{N+n} E_\infty \tag{12}$$

and

$$\Delta E_s = \frac{n}{N+n} E_\infty. \tag{12'}$$

In cases where $N \gg n$, one can assume that

$$\Delta E_s = \frac{n}{N} E_\infty. \tag{13}$$

Formula (12) applies to "ideal" solutions: it holds good only at weak concentrations.

Calculations show that the influence of impurities becomes appreciable only for strong concentrations; for sea water the reduction in the buoyancy amounts to only 2%, while for rain and river water it plays no role whatsoever.

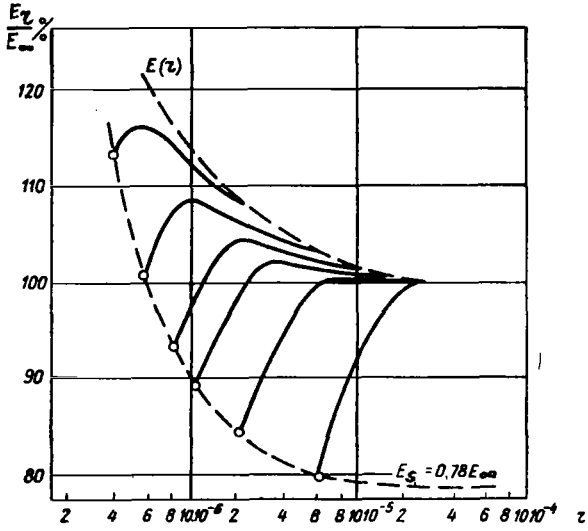


FIGURE 102. Buoyancy of vapor over droplets of NaCl solution as a function of droplet radius

Let us now turn again to the initial equation (3): $E = E_\infty + \Delta E_r - \Delta E_q - \Delta E_s$; introducing the expressions obtained for the individual terms we find

$$E = E_\infty \left[e^{\frac{c_r}{r}} - \frac{c_q}{r^4} - \frac{n}{N} \right] \tag{14}$$

or, approximately,

$$E = E_\infty \left[1 + \frac{c_r}{r} - \frac{c_q}{r^4} - \frac{n}{N} \right]. \tag{15}$$

For cloud droplets with $r > 10^{-6}$, as we saw, one can disregard the influence of electric charges. Then for such droplets we have

$$E = E_\infty \left[1 + \frac{c_r}{r} - \frac{n}{N} \right]. \tag{16}$$

This expression makes it possible to clarify a number of problems related to the condensation of water vapor in the atmosphere. Let us

assume that condensation takes place upon nuclei and that these condensation nuclei are hygroscopic and soluble in water. Then at the initial stage the embryonic droplet formed on these nuclei will be a saturated solution of this substance. Condensation nuclei frequently consist of particles of NaCl, for the saturated solution of which $E_s = 0.78\, E_\infty$.

The ratio $\frac{E_s}{E_\infty}$ (in percents) for a drop of saturated NaCl solution is represented graphically by the lower dashed curve in Figure 102. We will denote this function by $f(r)$. The upper dashed curve on this figures gives the values of $\frac{E}{E_\infty}$ for wettable and insoluble particles.

It is seen that for $r > 10^{-6}$ cm condensation on NaCl nuclei can begin at humidities even considerably lower than 100%. It may also be deduced from the graph that large insoluble but wettable particles are more active centers of condensation than very small hygroscopic nuclei ($r < 10^{-6}$ cm). Hence it follows that nonhygroscopic wettable particles, provided they are sufficiently large ($r > 1\ \mu$), can also act as condensation nuclei.

§ 3. Formation and growth of embryonic droplets

At the low supersaturations (less than 1%) prevailing in the atmosphere, spontaneous transition of water vapor into a liquid, as we saw, practically plays no role; minute embryos do form but are also instantaneously disrupted.

Condensation of water vapor on light ions of radius 10^{-8}–10^{-7} cm does not occur either, and only a few of the largest, the so-called ultra-heavy ions ($r > 10^{-6}$ cm), can initiate the formation of viable, growing embryonic droplets.

Condensation on nuclei is the principal element in the formation of embryonic droplets in the atmosphere.

From Kelvin's formula (cf. equation (5)) it follows that, for a given supersaturation, droplets can survive and may grow starting from a radius which we will denote by r_{em}:

$$r_{em} = \frac{c_r}{\ln \frac{E}{E_\infty}}. \tag{17}$$

If $r < r_{em}$ then, obviously, the droplets will evaporate. This condition holds for nonhygroscopic particles. If the nuclei are hygroscopic then, from equation (14), we find that

$$(r_{em})_{hygr} = \frac{c_r}{\ln \frac{E}{E_\infty} + \frac{n}{N}}. \tag{18}$$

Comparing these two expressions we see that in the case of hygroscopic particles the equilibrium radius $(r_{em})_{hygr} < r_{em}$, since the denominator of the expression for $(r_{em})_{hygr}$ contains the molar concentration of the solution $\frac{n}{N} > 0$ as an addend.

The value of $\frac{n}{N}$ depends on the concentration of the solution and decreases

as the droplet grows. Therefore when considering the process of formation and growth of embryonic droplets one must allow for this very important circumstance. Let us assume that at the initial instant we have a particle of salt of mass m_0. As molecules of water vapor settle on its surface part of the salt dissolves and the particle becomes covered by a film of water consisting of the saturated solution of this substance; a drop of saturated solution of radius r_{sat} is formed as a result. On further growth the mass of substance dissolved in the droplet remains constant (i. e., $m_0 = \text{const}$) while the mass of water increases in proportion to the volume of the droplet; consequently, the concentration of the solution decreases in inverse proportion to the volume, i. e., to the cube of the radius of the growing droplet. Therefore the ratio of the concentrations of the solution for two values of the droplet radius r_1 and r_2 is

$$\left(\frac{n}{N+n}\right)_2 : \left(\frac{n}{N+n}\right)_1 = r_1^3 : r_2^3$$

or, bearing in mind that $N \gg n$,

$$\left(\frac{n}{N}\right)_2 : \left(\frac{n}{N}\right)_1 = r_1^3 : r_2^3. \tag{19}$$

We will take the initial droplet size to be that of the saturated solution, i. e., $r_1 = r_{sat}$; then from (13) we obtain

$$\left(\frac{n}{N}\right)_1 = \left(\frac{n}{N}\right)_{sat} = \frac{(\Delta E_s)_{sat}}{E_\infty}. \tag{20}$$

As the droplet grows its radius $r = r_2$ at any instant, from (19), is given by

$$\left(\frac{n}{N}\right)_2 = \left(\frac{n}{N}\right)_1 \frac{r_{sat}^3}{r_2^3} = \frac{(\Delta E_s)_{sat}}{E_\infty} \frac{r_{sat}^3}{r_2^3} = c_s \frac{r_{sat}^3}{r^3}. \tag{21}$$

The value of the coefficient $c_s = \frac{(\Delta E_s)_{sat}}{E_\infty}$ can be determined reliably by experiment. Data for different salts are available in appropriate handbooks. Table 60 gives some of these values.

TABLE 60

Relative decrease in buoyancy of vapor for saturated solutions

	NH_4Cl	NaCl	$CaCl_2$	H_2SO_4 (50%)	H_2SO_4 (25%)
$\frac{(\Delta E_s)_{sat}}{E_\infty}$	0.20	0.22	0.65	0.68	0.15
$\frac{(E_s)_{sat}}{E_\infty}$	0.80	0.78	0.35	0.32	0.85

Introducing the value of $\frac{n}{N}$ from (21) into the approximate formula (16), we obtain

$$E = E_\infty \left[1 + \frac{c_r}{r} - c_s \frac{r_{sat}^3}{r^3}\right]. \tag{22}$$

For the droplets formed on particles of NaCl, for instance, on inserting the numerical values of the coefficients c_r and c_s we have

$$E = E_\infty \left[1 + \frac{1.2 \cdot 10^{-7}}{r} - 0.22 \frac{r_{sat}^3}{r^3}\right]. \tag{22'}$$

Analysis of (22) leads to the conclusion that at the initial instant, when the droplet is a saturated solution and its radius is r_{sat}, the buoyancy of vapor over the droplet can be less than E_∞ for condensation nuclei with $r > 5.5 \cdot 10^{-7}$ cm and condensation will begin at humidities below 100%. As the initial radius ($r = r_{sat}$) increases the relative humidity needed to initiate condensation decreases, but even for a plane surface ($r \to \infty$) it cannot be less than the ratio $\frac{(E_s)_{sat}}{E_\infty}$.

For NaCl this ratio is 0.78.

As the droplet grows and r increases the correction for concentration decreases more rapidly than the correction for curvature. In the case of NaCl it amounts to only $0.22/8 = 2.5\%$ for $r = 2r_{sat}$ and to less than 1% for $r > 3r_{sat}$.

Taking the derivative of (22) and setting it equal to zero we find that the function $\frac{E_s}{E_\infty}$ has a maximum at $r = \sqrt{3 \frac{c_s}{c_r} r_{sat}^3}$ and tends to $\frac{E_r}{E_\infty}$ upon further increase in r.

The family of curves $f(r)$ is represented in Figure 102 by solid lines. The curves were constructed for various initial sizes of the condensation nuclei. The ordinate of each point on these curves determines the relative humidity necessary for equilibrium for the given droplet size. Consequently, if the relative atmospheric humidity is greater than this value the droplets will grow, and if it is smaller they will evaporate. If the humidity is less than (say) 78% then condensation will not take place on particles of NaCl irrespective of the size of the nuclei.

Thus, for condensation of water vapor in the atmosphere it is necessary that the buoyancy (concentration) of vapor in the air be greater than that over the surface of the newly formed particles of the new phase and that the air contain minute particles capable of serving as centers of condensation.

However, these two conditions, though necessary for the initiation of condensation, are still insufficient for the formation of cloud droplets. The reason is that in unsaturated air only minute embryonic droplets can form on hygroscopic nuclei. Further growth of these droplets does not take place, for as the size of the droplets increases the concentration of the solution decreases very rapidly and the vapor buoyancy over their surface rises sharply. Therefore embryonic droplets formed on nuclei will be able to grow and transform into cloud droplets only if the buoyancy of vapor in the air exceeds that required for their appearance. As a rule this corresponds to a relative humidity $f = 101-102\%$, i.e., to a supersaturation $\Delta f = 1-2\%$.

As to the process of sublimation in the presence of ice crystals or frozen droplets supersaturation should be defined with respect to ice surfaces. Since the buoyancy of vapor over ice is lower than that over water, for large ice particles ($r \geq 10^{-4}$ cm) and sufficiently low sub-zero temperatures supersaturation over ice may even occur at relative humidities of less than 100% with respect to a plane surface of supercooled water.

§ 4. Condensation nuclei

From the preceding discussion it is evident that condensation nuclei play an important role in the initial stage of formation of droplets in the

atmosphere. It has been shown experimentally that particles of any finely dispersed substance can act as condensation nuclei, and in this respect it is difficult to draw the line between nuclei and atmospheric aerosols.

The number of condensation nuclei per cubic centimeter in the surface layer varies widely with local conditions — from several million in cities (especially industrial ones) to a few over the oceans. Vertically the number of nuclei in the troposphere decreases on the average by a factor of 10 as one rises over a height of the order of 2 km. The variation of the concentration of nuclei with height is expressed by the formulas cited in Chapter 2.

For illustration nuclei concentrations obtained during balloon flights are given in Table 61 (after Landsberg). A similar falling-off of the concentration was found from aircraft observations during the International Geophysical Year period (cf. Chapter 2, Figure 5).

TABLE 61

Average vertical distribution of nuclei

Height (m)	0–500	500–1000	1000–2000	2000–3000	3000–4000	4000–5000	5000
Concentration (cm^{-3})	22 800	11 000	2500	780	340	170	80

More or less considerable deviations from the data cited in the table and figure have been reported in individual cases depending on point of observation and local conditions. The concentration of nuclei increases under layers which inhibit turbulent transfer (inversion, isothermy), while inside clouds the concentration is lower than that in the absence of cloud. For example, the number of nuclei decreases by a factor of 2–3 in stratus clouds according to observations by Gaivoronskii (Moscow area) and by 30–40% in cumulus clouds according to V. A. Zaitsev (Leningrad area). Still, it appears that the number of nuclei in clouds remains considerable, and hence of all the nuclei only the most active participate in the formation of cloud droplets. Judging from the concentration of droplets in the lower section of clouds, the number of such active nuclei is 10^{-2}–$10^{3}\,cm^{-3}$; only for great vertical velocities (cumulus clouds) can condensation take place on less active nuclei as well.

Most condensation nuclei are so small that they cannot be seen with an ordinary microscope. It has been possible to determine their size only by indirect methods and, in recent years, by electron microscope measurements. At present condensation nuclei are classed into three groups according to size:

a) the smallest condensation nuclei, detectable in an Aitken counter and therefore called Aitken nuclei. Their radius is 5 10^{-7}–$2 \cdot 10^{-5}$ cm. This is the most numerous group of aerosols;

b) particles of aerosol with radius between $2 \cdot 10^{-5}$ and 10^{-4} cm, called large nuclei; their concentration in the atmosphere is 2–2.5 orders lower than that of Aitken nuclei;

c) particles with radius exceeding 10^{-4} cm, constituting the group of "giant" nuclei; their concentration is insignificant — several nuclei per liter of air.

Table 62 gives the mean concentrations and masses of the nuclei, after Junge.

The number of large nuclei on land, which averages 132 cm^{-3}, varies over a significantly narrower range than the concentration of Aitken nuclei, namely between 45 and 360: the number of giant nuclei is entirely negligible.

The mass of individual nuclei varies between 10^{-15} and 10^{-11} g, and that of giant nuclei reaches 10^{-8} g.

TABLE 62

Size distribution of nuclei, after Junge

	Radius (cm)					
	Aitken	large	giant			
	$5 \cdot 10^{-7}$–$2 \cdot 10^{-5}$	$2 \cdot 10^{-5}$–10^{-4}	$(1-2) \cdot 10^{-4}$	$(2-3) \cdot 10^{-4}$	$(3-5) \cdot 10^{-4}$	$(5-10) \cdot 10^{-4}$
Concentration (cm^{-3})	42 500	132	2.08	0.09	0.02	0.005
Mass (g/m^3)	17	25	23	4.2	5.1	9.1

The question of the nature and origin of condensation nuclei is very important. The nature (chemical composition) of condensation nuclei is usually studied via the chemical and spectral analysis of samples of raindrops and cloud droplets.

Results of the chemical analysis of rain water show that it contains on the average about 1 mg/l of Cl^- ions, 5–10 mg/l of SO_4^{2-}, 5–15 mg/l of HCO_3^-, 2 mg/l of Na^+, 3–5 mg/l of Ca^+ and other substances in smaller amounts. Maximum values tend to be 10–15 times greater and minimum values 10–20 times smaller than these averages.

In raindrops impurities may be captured during the fall and therefore results of analysis of the water obtained from cloud droplets shed more light on the chemical composition of nuclei. Though such data are still scanty, the presence of chlorides in amounts varying between tenths and several mg/l has been detected in all samples. A series of analyses of aerosol particles collected in various atmospheric layers has also been carried out lately (Junge et al.).

Thus one may say that the commonest nuclei are those containing compounds of chlorine, sulfur, nitrogen, carbon, magnesium, sodium, and calcium; nuclei consisting of NaCl are encountered especially frequently. The chemical composition of the nuclei varies to some extent from one area to the next. This is evidence that condensation nuclei enter the atmosphere from the earth's surface, though it is possible that a certain fraction is of cosmic origin.

To this day it is still unclear how many nuclei are delivered to the atmosphere by continental sources, and whether nuclei of maritime origin predominate. On land condensation nuclei appear as a result of rock weathering, volcanic eruptions and, chiefly, industrial activity. Over oceans nuclei are formed by the evaporation of sea spray, and also as a result of the direct evaporation of the components of sea salt. Certain authors (H. Köhler, R. I. Grabovskii et al.) have arrived at the conclusion that the

Pacific Ocean is the main source of nuclei (accounting for 80%) and that the process of atomization of sea water can completely supply the atmosphere with active condensation nuclei (chiefly chlorides). At the same time a number of observations indicate that a significant number of nuclei are of continental origin.

§ 5. Supercooled droplets. Formation of ice crystals

Supercooled droplets are observed in cloud and fog at temperatures down to $-30, -40°$; according to the observational data, moreover, the greater part of a cloud consists of liquid droplets down to temperatures of $-12, -16°$.

Until comparatively recently the presence of supercooled droplets in clouds was attributed to the presence in the droplets of dissolved substances. But since the concentration of solution in cloud droplets is of the order of 10^{-3} mole of dissolved substance per liter, a simple computation will show that the effect of impurities in lowering the freezing point of cloud and rain droplets is entirely negligible.

The many experimental studies carried out in recent years under laboratory and natural conditions show conclusively that the presence of supercooled droplets in the atmosphere can be explained by the conditions of phase transitions of water. For droplets to freeze it is necessary that a solid-phase embryo (center of crystallization) form in these droplets. Such centers of crystallization may be created by impurities contained in the droplets or may even enter them from the outside, but they may also form inside the liquid state.

The theory of phase transitions shows that the probability for the formation of a crystal embryo in a supercooled liquid as a result of density fluctuations is proportional to $e^{\frac{U+W_{em}}{kT}}$, where U is the free energy, W_{em} the work done in forming the equilibrium embryo, and k Boltzmann's constant.

Calculations show that the maximum probability for the formation of plane embryos occurs at temperatures of $-10, -12°$, and that for three-dimensional embryos at temperatures of $-50, -55°$. Thus very strong supercooling, and hence very low temperatures, are necessary for the formation of embryos. But the spontaneous formation of solid-phase embryos, as a probabilistic process, depends on the mass of cooled water; therefore small volumes of water (minute droplets) are more easily supercooled to lower temperatures than large volumes (large droplets). Thus, if there are no solid-phase embryos (ice crystals) in the atmosphere, freezing of the droplets, as numerous experiments have shown, will take place more rapidly as their size increases; this is why fine droplets supercooled to very low temperatures are encountered in the atmosphere.

The question of the primary formation of ice crystals in the atmosphere is much more complicated. It was long believed that the formation of crystals in the atmosphere took place in a manner analogous to condensation, through the sublimation of water vapor on nuclei; these were supposed to be different in nature from condensation nuclei and were named sublimation nuclei. But despite long searching they have never been detected in the atmosphere. The latest investigations show that their existence is not certain, since ice crystals can form either through the freezing of supercooled

droplets or, under appropriate temperature conditions, upon ordinary condensation nuclei and other aerosol particles.

Conditions of formation of ice crystals have recently been studied in laboratories and under natural conditions. Experiments carried out in special chambers (of the Wilson chamber type) showed that: droplets formed down to a temperature of −32°; individual crystals appeared at temperatures between −32 and −41°; cloud consisting exclusively of ice crystals formed only at temperatures below −41° Moreover— and this is very important— crystals appeared in the chamber when supersaturation with respect to a plane surface of water, and not ice, had been created in the chamber. From this it follows that in these experiments even at low temperatures (down to −41°) the liquid phase formed first and the solid phase (ice crystals) later.

Many successful experiments were also carried out recently in connection with the artificial stimulation of ice-crystal formation. It was found that the introduction of particles of solid carbon dioxide, at a temperature of −78°, in a system of supercooled droplets (to temperatures of −2, −5°) led to the appearance of an enormous number of crystals. Any substance cooled to temperatures below −39° will have the same effect. Strong cooling is apparently the main factor in these cases.

In another equally interesting series of experiments along these lines fumes of substances having a crystal structure similar to the lattice of ice were introduced into a fog consisting of supercooled droplets. The most active substances proved to be silver iodide and lead iodide, which belong to the hexagonal crystal system with a structure almost exactly like that of ice.

When these substances are introduced by volatilization into a supercooled cloud at temperatures below −5, −7°, an enormous amount of ice particles appears and within 5–10 min the entire cloud becomes crystalline. These results can be explained by assuming that 1) particles of silver iodide are artificial sublimation nuclei and the formation and growth of ice embryos begins on these nuclei, or 2) these particles, by settling on or being captured by supercooled droplets cause them to freeze, i.e., act as nuclei of crystallization. At present many facts indicate that, depending on conditions, the former process predominates in some cases and the latter in others.

Returning to the question of sublimation nuclei, we note the following. In nature sublimation nuclei include, in the first place, any minute particle of ice remaining in the atmosphere following the dissipation of a cloud or broken off from a larger dendritic crystal. Next, sublimation nuclei can consist of particles isomorphic to ice, such as the artificial AgI nuclei, which, however, have not been detected in nature so far. Finally, other soluble and insoluble particles may act as sublimation nuclei if they become coated with a film of water, which freezes at low temperatures; that is, at temperatures below −32° condensation nuclei may convert into so-called sublimation nuclei. It should be recalled that the buoyancy of saturated vapor over ice is lower than that over supercooled water. Therefore, when condensation nuclei and ice crystals (in particular, frozen droplets) of the same size occur and the [saturation] buoyancy of vapor over the hygroscopic particles (nuclei) is lower than over the ice crystals then, obviously, condensation will predominate; if the reverse is true (buoyancy lower over the ice crystals) sublimation will predominate.

Thus, depending on the size of the nuclei, their properties and temperature conditions in the atmosphere, either growth of ice crystals will take place or droplets will form. If there is a further decrease in temperature and crystallization nuclei are present, the latter may freeze. The experimental material available on crystallization nuclei suggests the following classification into three groups (after Mason):

The first group, consisting of nuclei effective at temperatures below −32°. These are rare under natural conditions. Artificial nuclei, namely iodides and other compounds (of which there are now more than 100), play a more important role.

The second group, consisting of nuclei appearing at temperatures between −32 and −41°; they also require supersaturation over water and should be classed among the freezing nuclei.

Finally, the third and more numerous group, consisting of nuclei active only at temperatures below −41°; their nature is relatively unknown.

In the presence of supersaturation embryos of ice crystals grow as a result of the diffusion of water vapor to their surface. Processes of two kinds play an essential role in the formation of the crystals: those associated with the internal mechanism of construction of the crystal lattice, and those reflecting the influence of the external environment.

For the construction of a regular crystal molecules must occupy a rigorously defined position in the crystal lattice. Therefore, of the molecules colliding with the surface of the crystal, far from all manage to find "their own place" Many molecules move — or migrate, as it is called — over the surface of the crystal for a certain period before being either fixed to it or evaporated.

Theoretical considerations put forward by Krastanow et al. show that molecules of water vapor will attach themselves more easily to an ice crystal at its vertices. Consequently, prisms, which grow in the direction of the principal axis, are the "ideal" form for ice crystals, and all other intermediate forms depend on external conditions.

If the number of molecules with which the crystal grows is small it will be able to build a regular, "ideal" crystal. If the number of molecules increases certain of these may be retained on the crystal surface and initiate the development of a new crystal lattice. This provides an explanation, albeit a qualitative one, for the appearance of different forms of ice and snow crystals.

Prisms form at low supersaturations (over ice) and low temperatures. Plates appear at high supersaturations, and dendritic crystal forms shaped as stars in the temperature region around −12°. The crystal form changes with changing external conditions — prisms are transformed into plates and, vice versa, plates into prisms or stars.

Chapter 19

FOGS AND CLOUDS

§ 1. General information on fogs and clouds

Condensation and sublimation of water vapor in atmospheric air lead to the formation of minute water droplets and ice crystals. Aggregates of such particles occurring directly above the ground are called fog; however, if the process takes place at a certain altitude in the atmosphere, clouds are formed.

From the physical standpoint fogs and clouds are colloidal systems consisting of water droplets, ice crystals and sometimes both together, falling slowly— or "suspended"— in the air.

As water droplet and ice crystals contained in air reduce its transparency, visibility in fog and clouds may be very poor. This optical property is used to ascertain the presence of fog and other forms of atmospheric turbidity. In fog (symbol ≡) daytime horizontal visibility usually drops below 1 km. If the visibility is greater than 1 km but less than 10 km,, which happens at low particle concentrations, this phenomenon is called mist (symbol =).

Occasionally deterioration of visibility is due to the presence in the air of solid particles of dust, smoke, and so forth, while the relative humidity remains low. This phenomenon is called haze (symbol ∞). Visibility in haze. as in mist, is less than 10 km.

Depending on their intensity as expressed by visibility conditions, the following types of fog and mist are recorded:

Thick fog (≡²)	Visibility	< 50 m
Moderate fog (≡)	"	50–500 m
Thin fog (≡°)	"	500–1000 m
Moderate mist (=, ∞)	"	1–2 km
Thin mist (=°, ∞°)	"	2–10 km

Fog is very unfavorable for aviation and land transportation and is also very common. The study of conditions of fog formation, and also of the geographical distribution of fog, is therefore of great practical importance.

The most general feature in the global distribution of fogs is that their frequency increases towards the high latitudes.

In polar regions fog is a frequent and persistent phenomenon; in the Arctic up to 100 days with fog are reported per year. In tropical latitudes fog is considerably rarer, with the exception of the humid subtropics and

coastal zones, where the number of days with fog reaches 60 per year. Inland the frequency of fogs decreases (due to a reduction in advection fogs). Near seas and oceans, and especially on raised coasts and shores lapped by cold currents, fogs become more frequent. For example, along the west coast of North America the fog incidence reaches 24–28 days in certain months, and up to 60–80 days with fog are recorded per year. Fog frequency is also high along the seaboard of the Far East districts of the USSR: about 80 days of fog in the year at Vladivostok, and many fogs (up to 85 days) on Dickson Island.

Fogs have a clear-cut daily and annual march. The daily march is particularly well-defined in the case of radiation fogs: these develop at night, reach maximum intensity by morning (at the time of the minimum temperature and maximum relative humidity) and dissipate after sunrise. Other types of fog are intensified by radiation processes. As a result the total fog frequency also displays a maximum at night and minimum by day (near the temperature maximum). The annual march depends on geographic conditions. In middle latitudes two maxima are observed in the fog frequency: one in autumn and the other in spring; over continents fog is commonest in autumn and over seas and oceans in spring, when the water surface is coolest.

The appearance, development and dissipation of clouds and fogs depend on general thermodynamic conditions in the atmosphere (macroprocesses), which determine the possibilities of condensation (sublimation) of water vapor: at the same time microprocesses inside clouds, which are due to special features of the structure of the cloud particles, are no less important and may sometimes be decisive.

In nature macro- and microprocesses operate simultaneously; usually it is only the final result of their joint action that is observed. A simultaneous examination of both would be a very complicated task.

We will first consider the results of the study of the microstructure of clouds and fogs and then examine the macroprocesses which govern their formation.

The fundamental and most important microphysical characteristics of clouds and fogs are:

1) the aggregate state of the particles;

2) the predominant size and spectra of the droplets, the form and size of the solid particles;

3) the volume concentration, i. e., number of particles per unit volume;

4) the concentration of condensed moisture per unit volume, or so-called liquid-water content of clouds and fogs. We note that the sum of the liquid-water content and gaseous moisture in a cloud gives the total moisture content of the air.

The complex of characteristics cited above is designated by the term "microstructure of clouds and fogs". In addition to these fundamental characteristics, the electric charges of the cloud particles, their chemical composition, radioactivity, etc. are also studied. To characterize the state of a cloud one also requires data on the temperature and humidity of the air in the cloud, vertical velocities of air movements inside it, intensity of turbulent mixing and so forth.

In the following sections we will consider the most important of the above-mentioned characteristics.

§ 2. Aggregate state of cloud particles

Above 0° all clouds and fogs consist of water droplets; melting snowflakes or hailstones fallen from overlying layers will be seen inside them only occasionally. As temperatures drop below 0° both supercooled droplets and ice crystals begin to appear. Finally, at very low temperatures (−40, −50°) clouds consist only of the solid phase of water— ice crystals in various forms.

Thus all clouds and fogs may be divided into three groups:

1) droplet (or liquid) clouds, consisting of droplets of water. The so-called supercooled clouds and fogs, composed of unfrozen droplets at $t < 0°$, occupy an important place in this group;

2) ice-crystal, or ice clouds, consisting entirely of ice crystals;

3) clouds and fogs of mixed structure, comprising both the liquid and solid phases of water (supercooled droplets and ice crystals).

According to the data available, it may be assumed that droplet fog consisting of supercooled droplets predominates down to very low temperatures (−20, −25°). At temperatures below −20° mixed fog is frequently observed; however, it is generally unstable and crystallizes rapidly when an ice phase appears inside it, which often falls afterwards. The relative amount of ice crystals usually increases with decreasing temperature but cases where supercooled droplets persisted down to −40° have been reported repeatedly. Below −40° fogs are usually composed of ice crystals.

More extensive data are available on the aggregate state of clouds. Observations show that in middle latitudes the three groups mentioned above occur with nearly equal frequency, pure droplet clouds having a maximum incidence in summer and minimum incidence in winter. According to E. G. Zak, out of 686 cases droplet clouds amounted to 34%, ice-crystal clouds to 37% and mixed clouds to the remaining 29% of all cases.

Droplet clouds without supercooled droplets always occur above the zero isotherm, while those containing supercooled droplets are mostly observed down to temperatures of −12°. On further lowering of the temperature supercooled droplet clouds become less frequent than mixed and ice-crystal clouds. Pure ice-crystal clouds, like ice fogs, are observed at temperatures below −40°. As to mixed clouds, sometimes the solid phase (ice crystals or frozen droplets) is confined to the upper region of the cloud, the lower parts consisting of the lqiuid phase. Alternatively, all the cloud or part of it may consist of a mixture of ice crystals and supercooled droplets. Finally, in some cases a layer with supercooled droplets lies above the ice-crystal part of the cloud.

Pure ice-crystal clouds include the high clouds (Ci, Cs, Cc). In the warm season low clouds (St, Sc), practically all subtypes of Cu (Cu hum, Cu cong), and the lower parts of Cb are entirely composed of droplets. In the cold season Ns–As and Cb are as a rule mixed; low clouds contain snowflakes and ice crystals together with minute droplets, and may thus be classed among mixed clouds, like the middle clouds (As and part of Ac).

§ 3. Microstructure of droplet fogs and clouds

Before turning to results of observations relating to the microstructure of fog and cloud, we present some information concerning the methods by which these results were obtained.

The basic method for determining the size of cloud particles is microphotography. Droplets are collected by various means on glass plates coated with a thin layer of oil or on celluloid with some other coating (soot, magnesium oxide, etc.) and the resulting "sample" photographed through a microscope. If one knows the magnification one can find the size of the particles. Figure 103 provides an example of a photograph of fog droplets.

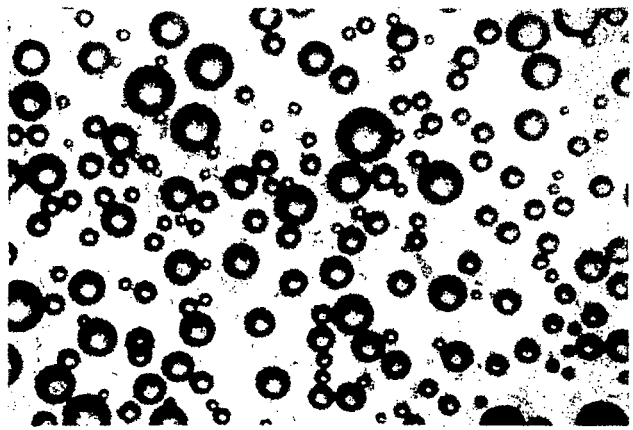


FIGURE 103. Microphotograph of cloud droplets

The method of microphotography permits reliable determination of droplet size from $r \geqslant 1\mu$ (resolving power of microscope) to $r \leqslant 60-70\mu$ (larger droplets may be deformed on capture). The fact that minute droplets flow past the obstacle during collection and are not captured must also be considered; hence the droplet spectrum obtained is distorted. To obtain the true distribution appropriate corrections must be introduced.

The simplest and most widely used method for determining the liquid-water content of clouds is due to V. A. Zaitsev. In his instrument air is sucked in at high speed together with cloud droplets through an intake nozzle; at a small distance from this nozzle is placed a filter paper previously coated with a dry dye. Droplets settling on the paper dissolve the dye and produce a colored stain. The liquid-water content (amount of liquid water, per unit volume, contained in cloud and fog droplets) can be found from the size of the stains by graduating the instrument in advance and considering the volume of air passing through. Other methods for determining the liquid-water content also exist; a description of these may be found in monographs by Zaitsev and A. A. Ledokhovich.

Direct measurement of the particle number is very complicated and is not usually performed; the practice is to compute this quantity from data on the liquid-water content and mean droplet size.

Microstructure of fog. Droplet fogs are encountered down to very low temperatures (−28° and lower). In most cases droplet sizes range from $r = 1\mu$ to $r = 35-40\mu$; only in isolated observations have droplets with $r \approx 60-100\mu$ been reported.

The range of droplet sizes and predominant droplet radii depend on conditions of fog formation. According to data collected by P. N. Krasikov, the range of droplet sizes narrows in evaporation fogs and at the initial stages of formation of other fog types (droplets from $r = 1\mu$ to $r = 20\mu$ and less predominate); the upper limit increases (to $r = 40-60\mu$) in the case of radiation and advection fogs. Prevalent sizes are: for evaporation fogs $r \approx 2-4\mu$, for persistent heavy fogs $r \approx 5-10\mu$, on the average $3-8\mu$ (the lower values are characteristic of thin fog at the initial stage of formation and the higher values of thick persistent fog).

While it is conceivable that fog and especially mist may also contain droplets smaller than $r \approx 1\mu$, they are not detectable by microphotography; their presence, however, is confirmed by optical data on the spectral transparency of fogs given by I. A. Khvostikov.

The liquid-water content of fogs, a_{liq}, amounts to $0.02-1.0\,g/m^3$ from direct observations and to as much as $1.7\,g/m^3$ according to results of calculations. Low values of a_{liq} ($a_{liq} = 0.02-0.3\,g/m^3$) apply to evaporation fogs at low temperatures (−10, −20°); high values ($a_{liq} = 0.3-0.9\,g/m^3$) to radiation and advection fogs. According to the observational data there is no clear relationship between the liquid-water content and type and temperature of formation of fogs, though in most cases the liquid-water content is lower at negative temperatures and greater at positive temperatures. On the average one may take $a_{liq} = 0.2\,g/m^3$.

The particle concentration, i. e., number of droplets per unit volume of fog, depends on the size of the droplets and the liquid-water content. On the average (from calculations) it amounts to 1−10 droplets per cubic centimeter for thin fog and 400−600 droplets per cubic centimeter for thick fog.

Microstructure of droplet clouds. Droplet clouds (without ice crystals) are recorded down to temperatures of −34, −36°. As temperatures fall below zero their incidence decreases; from 50% of all clouds at −8° to 10% at −20°.

The range of droplet sizes in these clouds is naturally greater than in fogs. The lower limit of the droplet radius (from microphotographs) is $r \sim 1-2\mu$, and the largest droplets, with $r \geqslant 100\mu$, are best classed among raindrops.

The smallest droplets (radius between 1−2 and $10-15\mu$; predominant radius between 3 and 4μ) occur in weakly developed cumuliform clouds (Cu hum) and sometimes in Sc and Ac. In St, Sc, Ns and Ac the droplet radius in most cases ranges from 1−2 to $20-22\mu$, with a predominant radius of $4-6\mu$. According to a few observations droplets with radius between 2 and 100μ are encountered in Cu, Cb and Ns.

Droplet size varies inside clouds with height above cloud base: as a rule minute droplets predominate in the lower and upper parts of clouds, and larger droplets in the central parts closer to the upper boundary.

On the average one can assume that droplets with radius between 3 and 6μ predominate in all clouds.

Liquid-water content of droplet clouds. Direct measurements by the Zaitsev and other instruments give values of a_{liq} varying between 0.01−0.02

and $1-4\,g/m^3$. However, zones with liquid-water contents exceeding $10\,g/m^3$ occur in individual cases in thick cumulus.

In St, Sc, Ns and As the liquid-water content is in most cases 0.05–$0.25\,g/m^3$ and comparatively seldom attains $1.2-1.5\,g/m^3$. In thick convective clouds — Cu cong and Cb — $a_{liq} \approx 1-2\,g/m^3$ (on the average). As one moves above cloud base the liquid-water content increases, later decreasing toward the cloud top. The vertical distribution of the liquid-water content in clouds depends on the stage of development or dissipation of the cloud, vertical variations in the velocity of updrafts, and so on. Recently G. K. Sulakvelidze and others discovered that large droplets may accumulate in convective clouds in the upper parts, creating zones of high liquid-water content ($a_{liq} > 10-20\,g/m^3$).

We pointed out earlier that concentrations in clouds depend on the liquid-water content and droplet size. Assuming that most cloud droplets have a radius of $4-6\,\mu$, for $a_{liq} = 0.1-2.0\,g/m^3$ their number $N \approx 10^2-10^3$ droplets/cm^3.

§ 4. Size distribution of droplets in fog and cloud

The size distribution of droplets can be represented in various ways. The simplest and clearest way is to construct diagrams showing the frequencies of droplets with various sizes; the droplet radius (or diameter) is plotted along the abscissa and the number of droplets N_i (or their fraction n_i %) having radii between r_i and $r_i + \Delta r$ along the ordinate; the interval Δr

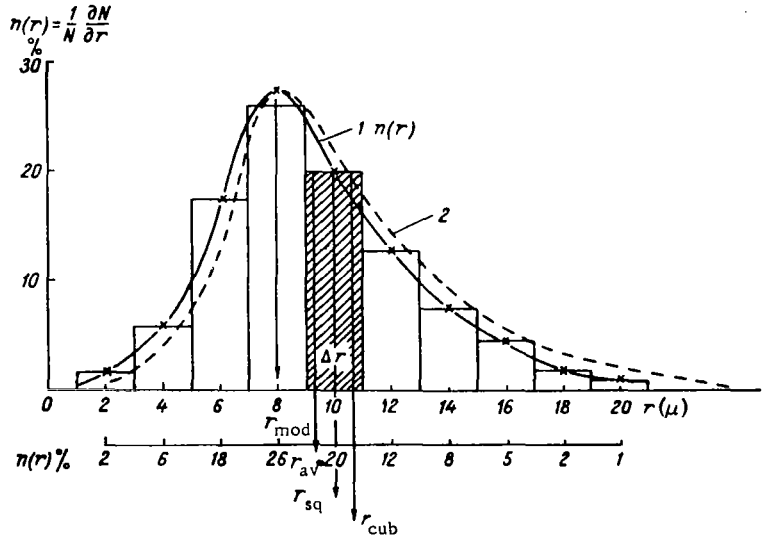


FIGURE 104. Typical size distribution of cloud droplets

1–from observational data; 2–from Khrgian-Mazin formula.

is kept constant. N_i changes at the transitions from one interval to the next and thus a step-like diagram is obtained (Figure 104). If N_i is referred

to the midpoint of the interval $\left(r_i + \frac{\Delta r}{2}\right)$, then, for large droplet numbers ($N \geq 10^3$), it is possible to obtain a smooth continuous "spectrum" of droplet sizes, i.e., the curve $N = f(r)$. Expressed analytically in the form of an equation, this dependence is called the distribution function (or curve; in Figure 104 it is shown by a dashed line).

In practice it is simpler to deal with the relative, rather than the absolute, number of droplets:

$$n_i = \frac{N_i}{N} \quad \text{or} \quad n_i = \frac{N_i}{N} \cdot 100, \tag{1}$$

where N_i is the number of droplets with the mean radius r_i and N the total number of measured droplets.

When the droplet radius changes by dr the relative number of droplets will change by the quantity

$$dn_i = \frac{1}{N} \frac{\partial N_i}{\partial r} dr = n(r)\, dr, \tag{2}$$

where $n(r) = \frac{dn_i}{dr}$ is the distribution function, which determines the fraction of droplets occurring in the range $dr \approx \Delta r$. On the graph (Figure 104) this fraction corresponds to the shaded area $n(r)\Delta r = n_i$. Obviously, the total number of droplets in this interval is

$$N_i = N n_i,$$

but since $\Sigma N_i = N$ it follows that the function $n(r)$ must satisfy the condition

$$\int_0^\infty n(r) dr = \frac{1}{N} \int_0^\infty \frac{\partial n}{\partial r} dr = 1 \text{ (or 100\%)}. \tag{3}$$

If the distribution function is expressed in terms of the absolute particle number then obviously

$$\int_0^\infty N_i(r)\, dr = N.$$

In addition to the curves of $n(r)$, which make it possible to determine the number of droplets with various sizes, in certain cases (e.g., when calculating the liquid-water content) it is necessary to know the relative distribution of droplets over their mass or volume, i.e., the relation $M = v\rho_d = f(r)$, where M is the mass, v the volume and ρ_d the density of the droplets.

Reasoning as in the derivation of the expression for $n(r)$, we find that the relative droplet mass $m_i = \frac{M_i}{M}$ occurring in the range with mean radius r_i varies with r according to

$$dm_i = \frac{1}{M} \frac{\partial M_i}{\partial r} dr = m(r)\, dr. \tag{4}$$

It is easy to pass over from the system of curves $n(r)$ to the system of curves $m(r)$. If in the range r_i the number of droplets were to increase or decrease by dN_i, the droplet mass in this interval would change by the quantity

$$dM_i = M_i dN_i,$$

but since $dM_i = Mm(r)dr$ and $dN_i = Nn(r)dr$,

$$M_i = \frac{dM_i}{dN_i} = \frac{Mm(r)\,dr}{Nn(r)\,dr},$$

whence

$$\frac{m(r)}{n(r)} = \frac{M_i N}{M} = \frac{\frac{4}{3}\pi\rho_d r_i^3 N}{\frac{4}{3}\pi\rho_d r_{cub}^3 N} = \frac{r_i^3}{r_{cub}^3}, \qquad (5)$$

where r_{cub} is the mean cubic droplet radius (for greater detail on this see below).

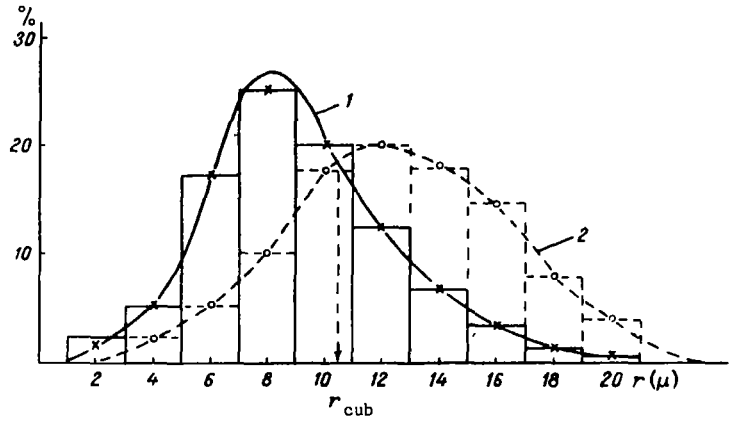


FIGURE 105. Distribution curves

1– $n(r)$; 2– $m(r)$.

Figure 105 gives the curves $n(r)$ and $m(r)$. Both are asymmetric; they rise steeply to a peak, then drop more smoothly toward the large droplets. The peak of the curve $m(r)$ is shifted to the right, since the liquid-water content is proportional to r^3 and therefore depends to a larger degree on the contribution of the large droplets.

Occasionally, especially in the literature outside the USSR, instead of the differential functions $n(r)$ and $m(r)$ so-called integral distribution functions are employed. These functions indicate what proportion of the particles (according to number or mass) has a radius greater or less than a specified value of r_i. This distribution can be obtained by integrating the function $n(r)$ from the specified r_i to ∞ and from zero to r_i. We set

$$F_1(r) = \int_{r_i}^{\infty} n(r)\,dr \quad \text{and} \quad F_2(r) = \int_{0}^{r_i} n(r)\,dr;$$

here the following condition must be satisfied:

$$F_1(r) + F_2(r) = 1 \ \text{(or 100\%)}.$$

The integral distribution curves are represented in Figure 106. The fractions of particles (in percents) are plotted along the ordinate and their radii along the abscissa.

Many formulas have been proposed in recent years for the analytic expression of the distribution function $N(r)$. Most of these can be reduced to

an equation containing four parameters, namely

$$N(r) = Ar^m e^{-br^k}. \qquad (6)$$

The coefficients A and b depend on given concrete conditions (liquid-water content of the cloud or fog, its age, etc.). The indices m and k play a more important role. Since the function r^m increases with increasing r, and the function e^{-r^k} decreases with increasing r, their product will give the form of the curve $N(r)$. The numerical values of the indices, according to Smolukhovskii and Schumann, are $m = 2$ and $k = 3$; according to Best $m = k_1 \approx 4$.

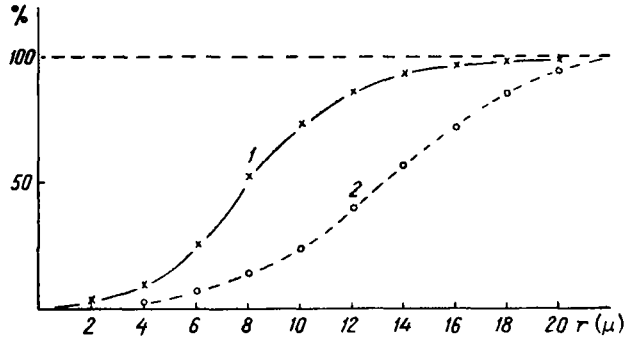


FIGURE 106. Integral distribution curves

1–n (r); 2–m (r).

The simplest and most convenient formula is that of Khrgian and Mazin, in which $m = 2$ and $k = 1$, i.e.,

$$N(r) = Ar^2 e^{-br} \quad \text{or} \quad n(r) = \frac{A}{N} r^2 e^{-br}. \qquad (7)$$

This formula gives a good description of the mean distribution of large droplets ($r > 2-4\mu$); in the small-droplet region ($r < 4\mu$) the values it produces are too low.

More exact formulas (A. N. Kolmogorov's normal-logarithmic distribution law and L. M. Levin's gamma-distribution) have so far failed to gain wide currency due to their complexity.

We now turn to a more precise definition of the concept of the mean droplet radius in fog and cloud. In the solution of many practical problems and for approximate calculations, the use of the droplet distribution spectrum proves to be excessively complicated. It would be desirable to obtain a mean particle radius such that the same particle number will produce the same effect as the actual distribution. Different means should be used depending on the problem under consideration. For example, the rate of change of the droplet mass upon evaporation and condensation is proportional to its radius r; the scattering of light on droplets is proportional to the square of the radius r^2; the change in their volume or mass is proportional to r^3; reflection of radio waves is proportional to r^6, and so on. In each such case one is required to determine the corresponding mean particle radius. We present expressions for a few of these.

1. Mean droplet radius

$$r_{av} = \frac{\Sigma r_i N_i}{\Sigma N_i} = \frac{1}{N}\int_0^\infty r N_i(r)\,dr. \qquad (8)$$

Since $\Sigma N_i = N$ (total number of droplets under consideration), it follows from (8) that

$$r_{av} N = \Sigma N_i r_i.$$

This means that if one takes N droplets of radius r_{av}, the effect they produce will be equivalent to $\Sigma N_i r_i$.

2. Modal droplet radius. This is the radius of the droplets most frequently encountered in the given system. It therefore corresponds to the peak of the distribution curve. Consequently, r_{mod} can be found from the condition

$$\frac{\partial N(r)}{\partial r} = 0.$$

3. Mean cubic droplet radius

$$r_{cub}^3 = \frac{1}{N}\sum r_i^3 N_i = \frac{1}{N}\int_0^\infty r^3 N_i(r)\,dr. \qquad (9)$$

Hence it follows that

$$\sum r_i^3 N_i = r_{cub}^3 N_i = \int_0^\infty r^3 N_i(r)\,dr.$$

Since the liquid-water content of the fog (cloud)

$$a_{liq} = \frac{4}{3}\pi\rho_d \int_0^\infty r^3 N_i(r)\,dr = \frac{4}{3}\pi\rho_d r_{cub}^3 N,$$

it follows that N droplets of radius r_{cub} will give the same mass of water as N droplets of different sizes for the given distribution. This example shows how much easier the calculation of the liquid-water content becomes when r_{cub} and the number of droplets per unit volume are known.

Other mean droplet radii can be determined in the same way: the root mean square radius r_{sq}, that of the droplets making the greatest contribution to the liquid-water content of the cloud r_m, and so forth.

In a homogeneous (monodispersed) cloud all particles are of the same size. For symmetric curves $r_{av} = r_{mod}$; in a polydispersed cloud

$$r_m > r_{cub} > r_{av} > r_{mod}$$

(cf. Figure 104).

Let us see, for instance, how the mean droplet radii are calculated from the Khrgian-Mazin formula $N_i(r) = Ar^2 e^{-br}$. *

1. The mean radius

$$r_{av} = \frac{1}{N}\sum r_i N_i = \frac{\int_0^\infty r^3 A e^{-br}\,dr}{\int_0^\infty A r^2 e^{-br}\,dr} = A\,\frac{6}{b^4}\,\frac{b^3}{2A} = \frac{3}{b}, \qquad (10)$$

whence $b = \frac{3}{r_{av}}$.

* Cf. the table of integrals $\left(\int_0^\infty r^k e^{-br}\,dr = \frac{k!}{b^{k+1}}\right)$

The typical forms of the droplet distribution curves for radiation and evaporation fogs are illustrated in Figure 107a and b. From these data it follows that at positive temperatures the spectrum in thick radiation fog tends to slope gently, spreading out in the direction of the large droplets, which reach $r \approx 60\mu$. At low temperatures in thin fog the spectrum becomes peaked; a peaked curve is also characteristic of the droplet spectrum in evaporation fogs.

The droplet spectra of various stratiform clouds are shown in Figure 108; Figure 109 gives the droplet spectrum of cumulus clouds.

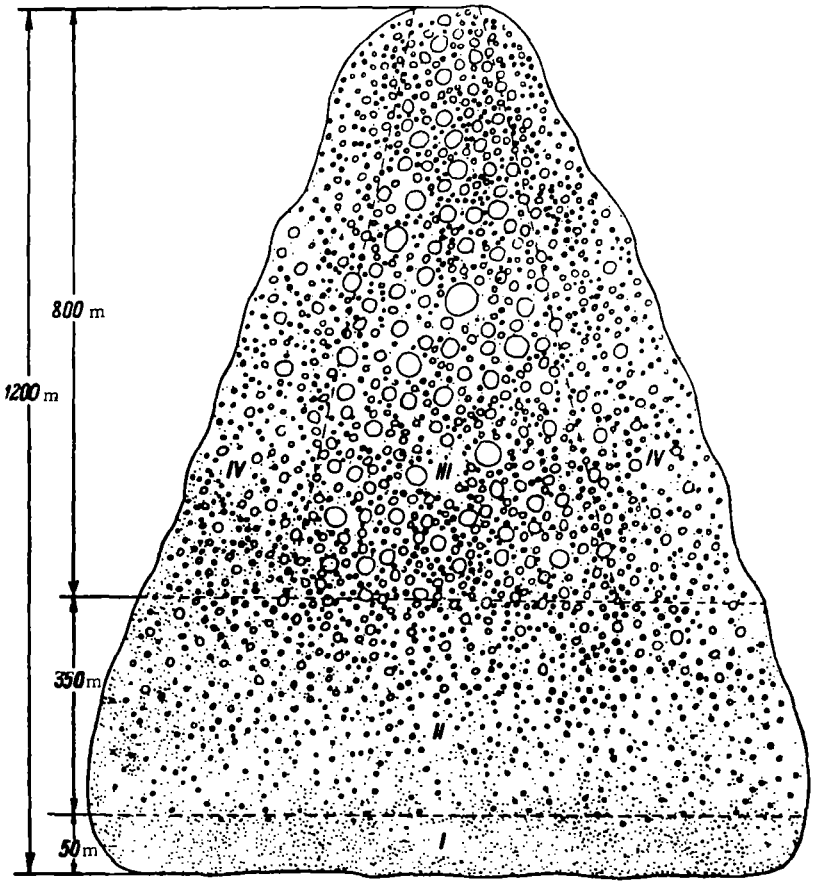


FIGURE 110. Scheme showing distribution of droplets in cumulus cloud (after V.A. Zaitsev)

It is important to note that the droplet spectrum is different in different parts of the same cloud. This variation is especially pronounced along the vertical. In clouds with considerable vertical speeds of air motion, as one moves above cloud base the mean droplet size increases gradually to a certain height. The most typical picture of such a vertical distribution was obtained in cumulus clouds as a result of work carried out at the Main Geophysical Observatory (1946–1948).

From Figures 109 and 110 it is seen that: at the base of cumulus clouds (zone I) droplets are fine and of very uniform size (peaked distribution curve); in the middle portion (zone II) the size of the predominant droplets increases and an appreciable number of larger droplets appears; in the upper section (zone III) the droplet distribution curve is even more deformed and indicates a general increase in the size of the cloud droplets (individual droplets reach $r = 100-200\mu$).

A similar variation of the droplet spectrum with height was also obtained for stratiform clouds (cf. Figure 108). However, in clouds of this type the largest droplets are sometimes recorded in the middle, and also in the lower, sections. This happens when ascending air motions are not decisive in the development of the cloud and the flux of water vapor from the upper warm, moist air is directed downwards.

The size of droplets near cloud edges is also influenced by the transport of water vapor into the surrounding medium and the evaporation of droplets; for this reason droplet sizes decrease at the tops of all clouds and also along the sides. Changes in the droplet distribution curve in different sections of cloud are determined by condensation and coalescence. The former tends to make the droplets more uniform in size; the latter results in a rapid increase in the number of large droplets and extension of the distribution spectrum toward the large droplets, the entire curve becoming smoother with a faint maximum.

It should be borne in mind that the above statements apply to average observational data. Individual clouds may display marked deviations according to the stage of development or dissipation of the cloud at which the droplet samples were taken.

§ 5. Microstructure of ice-crystal and mixed fog and cloud

The microstructure of ice-crystal and mixed clouds (fogs) is more complicated than that of droplet clouds and has not been studied as thoroughly. In ice-crystal clouds the individual particles of the solid phase of water differ in form and size. Mixed clouds contain frozen and unfrozen droplets of water, minute ice crystals of different forms and, occasionally, large ice formations (snowflakes, graupel, hail) which belong among particles of precipitation.

The basic method for studying particle sizes in such clouds remains microphotography. Instead of oil the glass is coated with a fast-drying lacquer; ice crystals colliding with this film of lacquer leave a trace ("replica"; cf. Figure 111). These replicas, which remain even after the crystals evaporate, are then processed.

The Zaitsev instrument, with suitable heating, is used to determine the liquid-water content of ice-crystal clouds.

The aggregate state of the system is determined visually and also by means of a phase-meter suggested by V. Ya. Nikandrov. In this instrument the flow of particles is scanned in a perpendicular beam of light; droplets produce a whitish veil while crystals are singled out by their bright sheen.

We present a few data on fog.

No rigorous temperature limit has been established for ice-crystal fogs; they have been reported from $-16°$ down. According to Klinov, pure

ice-crystal fogs always occur at temperatures below −40°. Mixed evaporation fogs have been reported by Krasikov between −11 and −28° and according to Wegener may be observed even at lower temperatures. These data

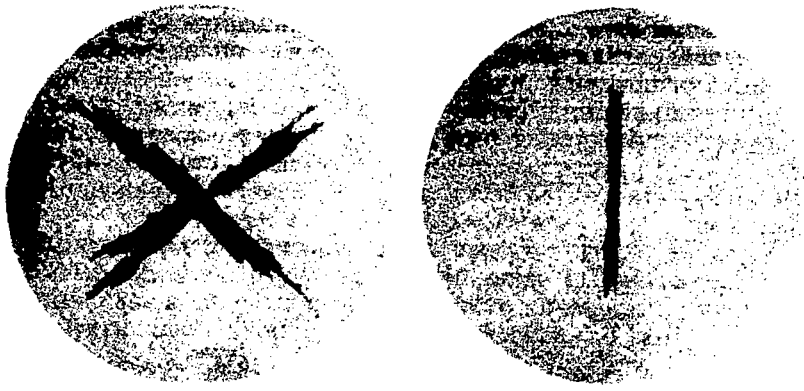


FIGURE 111. Microphotograph of "replicas" of ice crystals

indicate that droplets and crystals are very rarely observed in natural fog between 0 and −11° [sic]; between −10 and −16° mixed fogs are fairly frequent; and below −16° pure crystal fogs are also observed.

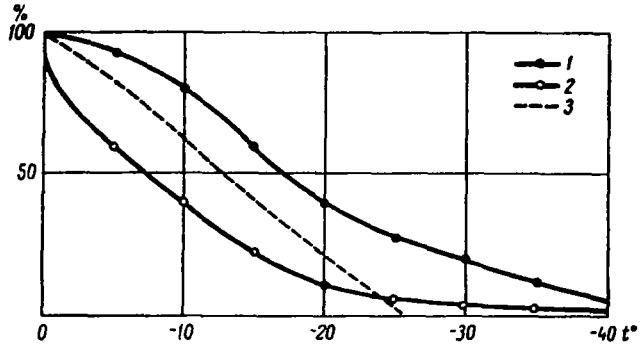


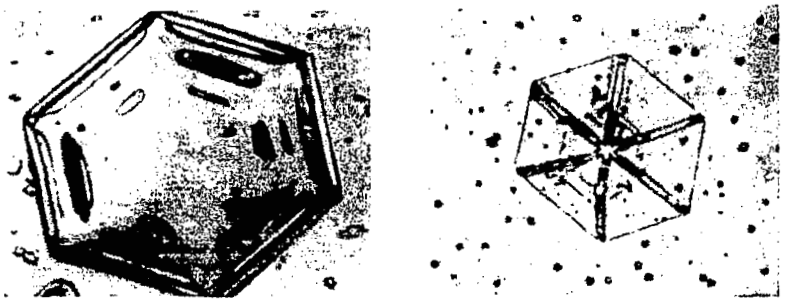
FIGURE 112. Percentage of droplet and mixed clouds

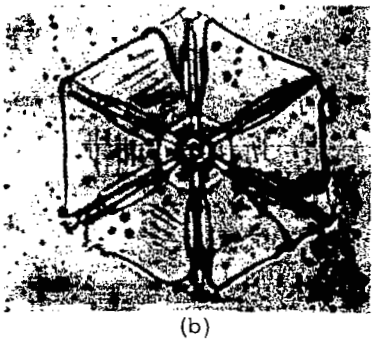
1-droplet and mixed clouds together; 2-pure droplet (after Borovikov and Sakhno); 3-droplet (after Peppler).

Typical crystal forms in ice-crystal and mixed fog are plates (thin prisms), stars and occasionally elongate needle-shaped prisms. The size (diameter) of plates and stars varies between 10–15μ and 500μ. According to Krasikov, the predominant plate size is roughly 25–55μ in light fogs, 35–75μ in moderate fogs and 25–200μ in thick fogs.

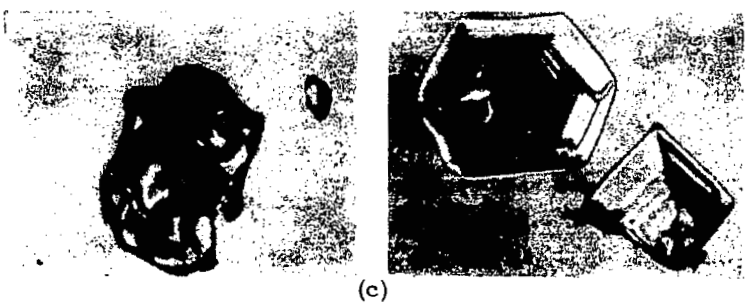
It is interesting to note that, according to Nikandrov, in mixed evaporation fogs the number of minute droplets ($r<3\mu$) drops sharply and the

(a)





(b)



(c)

FIGURE 113. Various crystal forms (after A. M. Borovikov)

larger droplets ($r = 5-6\mu$) become dominant. This is due to the fact that in the presence of crystals small droplets evaporate more rapidly than large ones.

The liquid-water content of ice-crystal and mixed fogs is low — in most cases $a_{liq} < 0.03\ g/m^3$ — and cannot be measured with a Zaitsev instrument. The concentration of crystals in ice-crystal fog is less than one per cubic centimeter. The ratio of crystals to droplets in mixed fog has not been evaluated quantitatively and may obviously vary from 0 to ∞.

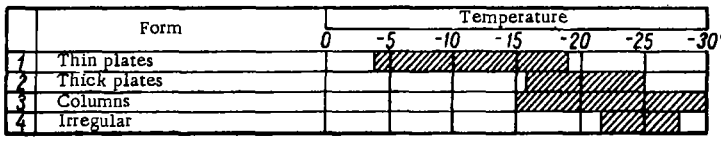


FIGURE 114. Crystal forms as a function of temperature in clouds

At present there is more observational material on the microstructure of ice-crystal and mixed clouds than on that of fogs. The relative number of droplet and mixed clouds in middle latitudes, according to A. M. Borovikov and L. G. Sakhno, is about 60% at −15° and about 5% at −40° (cf. Figure 112); that of ice-crystal clouds is respectively about 40 and 95%. Clouds in various aggregate states can be observed at any sub-zero temperature but the proportion of ice-crystal clouds increases with falling temperature while that of mixed clouds decreases.

The diverse crystal forms encountered in clouds, the most important of which is the hexagonal prism, can be reduced to three groups. The crystals of the first group grow along the principal axis and comprise elongate, occasionally hollow prisms shaped as "needles", "columns" and their combinations (Figure 113a). Crystals in the second group grow along

TABLE 63

Crystal form	Temperature (°)	Crystal size (μ)
Thin hexagonal plates	From 0 to −15	From 50 to 500 (diameter) from 10 to 20 (thickness)
Star-shaped crystals showing dendritic structure		From 0.5 to $5 \cdot 10^3$ (diameter)
Thick hexagonal plates		200 (diameter)
Prismatic columns (single prisms and groups)	From −15 to −30	200 (length)
Clusters of prismatic columns		$1 \cdot 10^3$ (diameter)
Hollow prisms	Below −30	$0.5 \cdot 10^3$ (length)
Single hollow prisms		100 (length) 1–5 (diameter)

secondary axes to form more or less regular hexagonal plates and stars (Figure 113b). The third group includes crystals of indeterminate or irregular shape, "pyramids", "studs" and so on (Figure 113c).

Borovikov cites the following temperature intervals for the occurrence of individual groups: thin plates — 0 to −16°; thick plates — from −16 to −25°; prismatic columns — from −13 to −30°; irregular forms — from −10 to −28° and below (Figure 114).

These results agree fairly well with Wall's scheme for the growth of crystals in supercooled fogs.

The dimensions of ice crystals vary widely and depend on conditions and duration of growth. Needles are several microns in diameter and up to one millimeter or more long; plates are one micron thick and over a millimeter in diameter. Not only temperatures but also the supersaturation plays a role in the formation of crystals. A study of crystal forms in different clouds led Weickmann to the generalization described in Table 63.

Judging from indirect data the liquid-water content of ice-crystal clouds is very limited, namely less than 0.02 g/m^3; that of mixed clouds is greater — up to 0.2–0.3 g/m^3.

In ice-crystal clouds, in view of the low liquid-water content and large size of the particles, the concentration of crystals amounts to less than 1–2 crystals per liter. In mixed clouds the concentration of droplets is of the order of $10–10^2$ per cubic centimeter and that of crystals is low.

§ 6. Processes leading to the formation of fog. Classification of fogs

In the presence of fog the moisture in the air consists of water vapor and liquid or crystalline water. Consequently, the moisture content per unit volume (g/m^3) can be expressed as follows:

$$W = a_{vap} + a_c, \tag{17}$$

where W is the total moisture content of the air in the presence of fog, a_{vap} is the vaporized moisture and a_c is the condensed moisture.

From (9) it follows that $W > a_{vap}$, i.e., for the formation of fog the total moisture content of the air W **must become** greater than the amount required for saturation at the given temperature. This can happen under the following conditions.

1. The moisture content W does not change but the content of water vapor a_{vap} decreases due to cooling of the air. Lowering of the temperature can result from: a) radiative cooling, b) heat exchange with the earth's surface and surrounding air masses, and c) adiabatic expansion of the air mass upon vertical ascent.

2. W increases due to evaporation while a_{vap} remains constant. This happens, for instance, when cool air comes into contact with a warmer evaporating surface.

3. Both W and a_{vap} change; this can happen when two air layers with different temperatures and moisture contents are mixed together.

In nature fogs are mostly formed through the simultaneous operation of all three factors, and the relative importance of each factor depends largely on weather (synoptic) conditions and local features. Radiative cooling, for example, undoubtedly assumes greater importance in clear weather than in overcast weather; evaporation is stronger over moist surfaces (in particular basins of water) than over dry ones (deserts, etc.); mixing

(horizontal and vertical) over these surfaces varies and depends moreover on local relief.

The fact that fogs are caused by a complex aggregate of simultaneously acting factors results in wide variety and makes it very difficult to classify them.

Fogs are classed as air-mass and frontal according to synoptic conditions; this, however, is a very general subdivision, and in each of these groups the development of the fog may be due to different factors.

As noted earlier, fogs may be classed in the following groups according to the physical causes involved: evaporation fogs, cooling fogs and mixing fogs

1. Evaporation fogs. These are due to the influx of water vapor into the air upon evaporation from a warm underlying surface and subsequent chilling of the air rising over the surface; for this it is necessary that the evaporating surface be appreciably warmer than the air. One can distinguish between:

a) sea fogs, observed as a rule over unfrozen bays and ice clearings in winter at sufficiently low air temperatures. These are particularly intensive in polar regions;

b) autumn fogs over rivers and lakes, which form in autumn when the water surface in rivers and lakes is appreciably warmer than the ground. When cooler land air advances over the water intensive evaporation often results in fog.

2. Cooling fogs. These include the commonest and also the most intensive fogs. They are due to chilling of air and are classed into three types according to the causes responsible for the chilling:

a) radiation fogs, due to cooling of the surface and adjoining layer of air by radiation and turbulent mixing. The surface is cooled mainly by radiation but the air loses little heat by this process; its chilling is due chiefly to turbulent heat transfer between the surface layer and ground. Condensation occurs in the air when the air temperature drops below the dew point. Radiation fogs occur in the evening and at night; favorable conditions include the absence of a cloud cover, high relative humidity and low wind speeds;

b) advection fogs, formed in relatively warm air masses advancing over cooler sections of the surface. When a warm air mass is transported over a cold surface an inversion is established inside it. The inversion gradually envelopes greater layers of air. Condensation of water vapor begins, like the chilling, from the surface and may spread over the entire inversion layer, which sometimes extends to considerable altitudes (1.5–2 km). Advection fogs are particularly common along sea coasts in the cold season due to air advancing over land from the warm surface of the water (so-called coastal fogs). They are also frequent over cool ocean currents.

c) hillside fogs (fogs of ascent), formed as a result of adiabatic cooling when air rises along hillsides and hence sometimes called orographic fogs.

3. Mixing fogs. These are due to horizontal mixing of air masses having different temperatures and humidities and also, under certain conditions, to turbulent vertical transfer. Let us consider the effect of mixing. We will confine ourselves to the simplest case in which two air masses of equal sizes are mixed, their temperatures are t_1 and t_2 and each contains a quantity of water vapor in the state of saturation (buoyancies E_1 and E_2).

As a result of mixing the mean temperature of the mixture should be $t_{av}=\frac{t_1+t_2}{2}$ and the buoyancy of the vapor $E_{av}=\frac{E_1+E_2}{2}$; but this, from Figure 115, would be greater than the buoyancy E corresponding to t_{av}. An amount of water vapor corresponding to the difference $E_{av}-E$ should condense and form a fog.

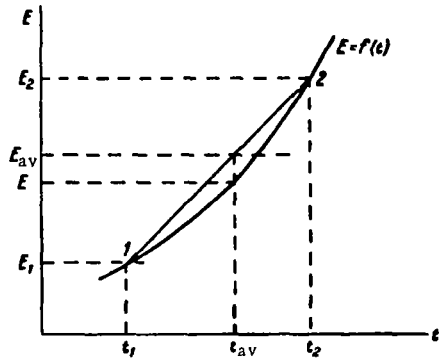


FIGURE 115

Calculations show that the effect of mixing will be felt only in the case where the humidity of the air masses in question is not less than 95% and the temperature difference between them not less than 10°. Therefore such fogs are usually of limited intensity.

Fogs traceable to human interference occupy a rather special position. These fogs are primarily urban; they are a frequent phenomenon over large cities, where an abundance of active condensation nuclei of industrial origin favors the onset of condensation in moist air. They are characterized by great intensity and long duration.

Such fogs may arise from any of the causes listed above but they are aggravated by the influence of urban conditions and may sometimes persist for several days. To this group belongs the special type of fog formed over populated points of Siberia in the stove-heating season (but mostly during severe frosts), which is due to the large quantities of nuclei and water vapor released into the air in the combustion of firewood and coal.

§ 7. Physical foundations of fog forecasting

Since the impaired visibility in fog affects the normal functioning of all forms of transportation a great deal of attention has been devoted to the development of methods of fog forecasting. Without going into synoptic methods which involve computation of air-mass movements — an absolutely essential point in the case of advection fogs — we note that other types of fog can be predicted by evaluating the basic factors responsible for their appearance. Most works deal with physical methods for forecasting radiation and, of late, evaporation fogs. We present certain of the thermodynamic relations on which these methods are based.

For the formation of fog it is obviously necessary that the humidity of the air reach saturation. Let us consider the conditions under which this happens. We take the logarithmic derivative of the expression for the relative humidity $f = 100\frac{e}{E}$

$$\frac{df}{f} = \frac{de}{e} - \frac{dE}{E}. \tag{18}$$

Introducing the value of dE from the Clausius-Clapeyron equation $\frac{dE}{E} = \frac{L dT}{AR_{vap}T^2}$ into the above, we obtain

$$\frac{df}{f} = \frac{de}{e} - \frac{L}{AR_{vap}T}\frac{dT}{T}. \tag{19}$$

From this one sees that the growth of the relative humidity and the attainment of the state of saturation are governed by the increase of the buoyancy e and decrease of T $(dT < 0)$. Under conditions of formation of radiation fogs $\frac{de}{e}$ is very small and the main term is the second one in the right-hand side of equation (19), which is 10–20 times larger than $\frac{de}{e}$ (since under average conditions the factor $\frac{L}{AR_{vap}T}$ is about 19.5). Consequently for this type of fog one can take $\frac{de}{e} = 0$. Then, integrating equation (19) over f from $f = f$ to $f = 100\,\%$ and over T from $T = T$ to $T = \tau$ (where τ is the dew point), we obtain

$$\ln\frac{100}{f} = -C\ln\frac{\tau}{T}, \tag{20}$$

where $C = \frac{L}{AR_{vap}T}$.

But $\tau = T - \Delta T_1$, where ΔT_1 is the required lowering of the temperature to the dew point, and therefore

$$\ln\frac{\tau}{T} = \ln\frac{T-\Delta T_1}{T} = \ln\left(1 - \frac{\Delta T_1}{T}\right) \approx -\frac{\Delta T_1}{T};$$

consequently,

$$\ln\frac{100}{f} = C\frac{\Delta T_1}{T}. \tag{21}$$

Hence, passing over to common logarithms, we obtain

$$\Delta T_1 = \frac{2.3T}{C}(2 - \lg f) \approx 0.115T(2 - \lg f), \tag{22}$$

which makes it possible to calculate ΔT_1 from the initial values of T and f.

In the case of radiation fogs data of observation in the evening hours are usually taken as initial values. The development of a fog still requires a further drop in temperature ΔT_2, upon which a certain quantity of water vapor Δa g/m^3, which determines the liquid-water content of the fog (a_{liq} g of water/m^3), will pass into the liquid state. This temperature drop can be found by introducing the expression for $\frac{de}{e} = \frac{dE}{E}$ from the Clausius-Clapeyron equation into the expression for the logarithmic derivative of the

absolute humidity ($a = 217\left(\frac{e}{T}\right)$ g/m^3) $\frac{da}{a} = \frac{de}{e} - \frac{dT}{T}$. Then

$$\frac{da}{a} = \frac{L}{AR_{vap}T}\frac{dT}{T} - \frac{dT}{T} = \left[\frac{L}{AR_{vap}T} - 1\right]\frac{dT}{T}. \quad (23)$$

Passing over to finite differences and recalling that in fog $T = \tau$, we find

$$\Delta a = a_{liq} = a_\tau\left[\frac{L}{AR_{vap}\tau} - 1\right]\frac{\Delta T_2}{\tau} = 217\,\frac{E_\tau}{\tau^2}\left[\frac{L}{AR_{vap}\tau} - 1\right]\Delta T_2. \quad (24)$$

The above expression allows one to compute the liquid-water content of a fog from the temperature drop ΔT_2 below the dew point, or to find the temperature drop required for the formation of a fog with a definite liquid-water content.

The value of the factor $217\frac{E_\tau}{\tau^2}\left[\frac{L}{AR_{vap}\tau} - 1\right]$, which is equal to the liquid-water content for $\Delta T_2 = 1$, is easy to compute. It depends largely on τ, increasing as the latter increases, from which it follows that fogs formed at higher temperatures have a greater liquid-water content. Thus if radiative lowering of the temperature by the amount ΔT has been forecast, from (22) and (24) it is possible not only to predict the onset of a fog but even to compute its liquid water content.

For other types of fog the quantity $\frac{de}{e}$ in equation (19) cannot be disregarded and under certain conditions becomes paramount. Such conditions obtain for evaporation fogs when the appearance and development of the fog are due to an increase in the moisture content of the air by evaporation from an open water surface. Taking $dT = 0$ in this case, we reduce equation (19) to the form

$$\frac{df}{f} = \frac{de}{e}. \quad (25)$$

From this it follows that the prediction of the fog reduces to prediction of the variation of the quantity $\frac{de}{e}$. M. P. Timofeev gives the following formula based on a theoretical consideration of this problem:

$$\frac{de}{e} = \frac{1}{1 + \frac{e_1}{E_{vap} - e_1}F}, \quad (26)$$

where e_1 is the humidity of air advancing over the evaporating surface, E_{vap} is the saturation vapor pressure over this surface and F is a function which depends on the size of the basin of water and the turbulent exchange between the water surface and air (the author constructs graphs for the determination of this function).

In the general case when forecasting evaporation and especially advection fogs one must consider both the variation of the moisture content of the air and the variation of its temperature. Solving this problem theoretically with certain simplifications, Timofeev estimated the values of the initial atmospheric humidity f_1 and temperature difference $T_{vap} - T_1$ between the underlying surface and air for which the formation of fog could be expected

§ 8. Basic processes responsible for the formation of clouds

The most important factor involved in the formation of clouds is chilling of humid air, which can happen due to the following causes:

1) adiabatic expansion of air on vertical ascent,
2) turbulent transfer,
3) radiation (radiative chilling).

Cooling of air during adiabatic expansion involves a reduction in pressure. The main factor here is the movement of air into higher atmospheric layers. Variations in pressure at any single level amount on the average to 5–6 mb per day and rarely exceed 15–20 mb even during the passage of fronts. Pressure drops of this kind can chill the air only by 1–2° per day. In vertical ascent, on the other hand, air containing unsaturated water vapor is cooled adiabatically by 1° for each 100 m of ascent.

When convection is well developed the air may rise by a height of the order of kilometers, experiencing very strong cooling in the process.

Chilling of air by turbulent transfer and mixing depends on the vertical distribution of temperature. In a stable stratification the upper portions of the layer in which turbulent transfer takes place will be cooled. If this cooling is accompanied by the transport of nearly saturated water vapor, its condensation may lead to the formation of stratus clouds (St).

Finally, the third cause of chilling is radiation. This process is manifested in the cooling of air layers containing large amounts of water vapor together with dust particles, condensation nuclei and smoke particles, and, in particular, in nighttime chilling of the upper cloud boundary. Radiation often results in the appearance; and sometimes intensification, of the comparatively thin nighttime sub-inversion clouds of the St and Sc type.

In nature all these processes operate together, overlapping on one another. However, the main factor in the formation of clouds, as noted above, is always the vertical movement of air; such movements, in fact, are the basic cause responsible for the formation of large cloud masses of considerable horizontal and vertical extension.

Depending on the character of the vertical air movements and other physical processes, clouds differ in both appearance and internal structure.

The largest cloud group comprises those due to large-scale ascending movements. These clouds extend over large surfaces in the horizontal direction and occupy a considerable portion of the atmosphere in the vertical. Conditions favoring the appearance of such vertical motions occur in frontal zones. Their vertical speed is limited (not more than a few cm/sec). The so-called frontal clouds formed in this process along the surface of discontinuity between warm and cold masses, which lies at a small angle (30–40') to the horizon, constitute an entire system which includes a wide variety of cloud types and subtypes.

Figure 116 illustrates a warm-front cloud system (after S. P. Khromov). As one can see, the advance of the warm front brings a sequence of stratiform clouds: cirrus, cirrostratus, altostratus and, lastly, the very thick nimbostratus and associated fractonimbus. Along the horizontal this system frequently extends over an area hundreds of kilometers (up to 1000 km) wide and several hundreds and thousands of kilometers long (along the front).

Figure 117 illustrates the cloud system formed at cold fronts, i.e., at fronts advancing towards the warm air mass, for two types of cold front (first and second order)

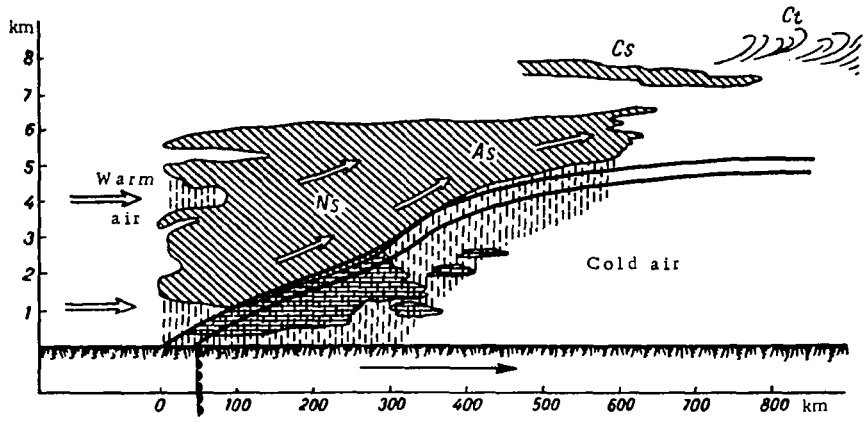


FIGURE 116. Warm-front cloud system (schematic)

As one can see, the cloud system formed here also consists of stratiform types; however, the sequence is reversed, and cloud forms begin with Ns and end with As-Cs. A characteristic feature of these fronts is the presence of thick Cb clouds ahead of them

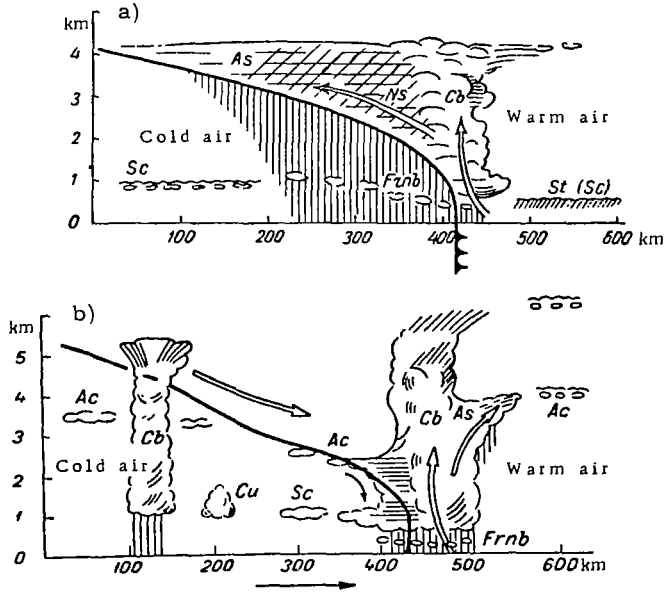


FIGURE 117. Cold-front cloud systems (schematic)

Without going into details we note that actual frontal cloud systems may differ considerably from the schemes shown in the figures, a circumstance to be attributed to variations in the properties of air masses and to dynamic factors.

The most intensive of all ascending air motions are associated with convection. As we noted earlier, convection can have two causes, thermal and dynamical.

Thermal convection is particularly intensive in the presence of unstable atmospheric stratification. Over land it appears in the warm season in daytime as a result of uneven heating of the surface (forest and steppe, tilled land and meadow, etc.). Over seas it appears mostly at night when the surface of the water is warmer than overlying layers of air. Warmed volumes of air may continue to rise until their temperature equals that of the surrounding medium. High instability will lead to great updrafts with vertical speeds of up to several meters per second.

Dynamic convection denotes vertical air motion occurring ahead of a cold-front slope or in zones of convergence of air currents. In such cases enormous masses of air are lifted over a height of several kilometers over a large area.

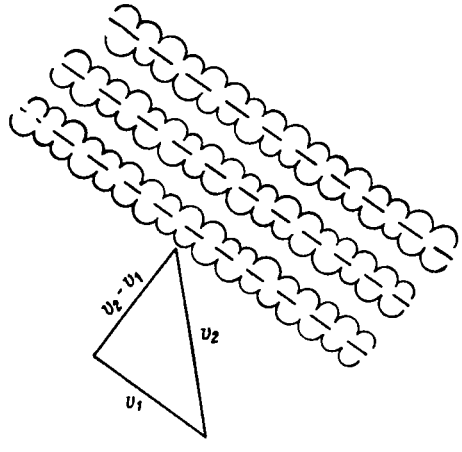


FIGURE 118. Vector difference of the velocities of two atmospheric layers and the direction of cloud banks

The velocity of vertical streams in regions of well-developed convection may be very considerable and varies over a broad range (from a few cm/sec to 10–20 m/sec). These vertical motions give rise to cumuliform (convective) clouds consisting of individual cloud masses strongly developed along the vertical for a comparatively limited horizontal extension. These include fair-weather cumulus, (Cu), thick cumulus (Cu cong), cumulonimbus and other forms.

Finally, a very extensive and variegated group comprises the so-called wave clouds, consisting of horizontal layers of waves, rolls, banks, etc.

Wave clouds are formed at all three levels (high— Cc, middle - Ac and low— Sc); their vertical extension is usually limited to a few tens of hundreds of meters.

It was long believed that all wave clouds were the visible manifestation of the gravitational "Helmholtz waves". At the end of the previous century Helmholtz demonstrated that if two streams of air in the atmosphere flow one above the other at different speeds (the lower one being denser and cooler than the upper one), waves may form at the surface of discontinuity between these layers. If the water vapor contained in the air is close to saturation, condensation and the formation of long cloud rolls may occur at the crests of the waves, where the ascending motion takes place; cloud-free spaces will remain where the descending motion occurs. If the vector velocities of these two streams $\vec{v}_1$ and $\vec{v}_2$ are directed at a certain angle to each other (Figure 118), the resulting waves will extend at right angles to the vector difference $\Delta\vec{v} = \vec{v}_1 - \vec{v}_2$.

However, as later research revealed, the formation of wave clouds may also be due to other processes. Indeed, when gravitational waves are formed in the atmosphere, as D. L. Laikhtman has shown, in most cases they become unstable and are destroyed. Their degree of stability is

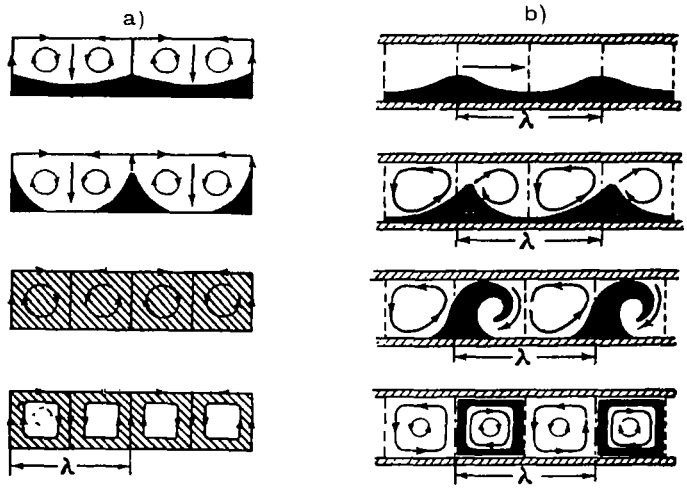


FIGURE 119. Diagram illustrating the gradual formation of longitudinal (a) and transverse (b) cells in a layer of moving air

governed by the difference in temperature and speed between the two streams under consideration: the smaller the temperature difference the lower the stability of the wave. For instance, for a temperature difference of 5.5° between the moving layers, all waves longer than 100 m will be unstable. The waves disintegrate and convert into a series of eddies rotating alternately in opposite directions. An entirely different form of motion, called "cellular circulation" (Figure 119), appears.

The theory of cellular circulation in the atmosphere was developed by Rayleigh, Jeffreys, N. S. Shishkin, L. S. Gandin and others. Their work makes it possible to determine the conditions under which loss of stability will occur in a layer, and also the dimensions of the resulting cells.

It is possible to demonstrate that cellular circulation develops in those

atmospheric layers in which the vertical temperature gradient exceeds 0.82°/100 m. When such circulation occurs in the atmosphere cloud will form in the parts of the cells where the air rises and clear spaces where descending currents occur. Clouds with "cellular" structure will form as a result (varieties of Sc, Ac and Cc).

Wave clouds may also appear when air flows over mountain ranges and other obstacles on the earth's surface. In this case a system of waves which retains its position in space is formed above and behind the range. The theory of such waves is given by A. A. Dorodnitsyn.

This gives one an idea of the wide variety of cloud forms that can appear in the atmosphere depending on the type of vertical motion involved and many other factors. Differences in cloud-forming processes are reflected not only in the diversity of appearance but also in some measure in the microphysical structure of clouds. The deficiencies of a purely morphological classification of clouds based on appearance and height thus becomes evident; genetic features must be considered as well. This point is taken into account to some extent in the present international classification, although in station observations the morphologic criteria remain paramount.

Detailed data on the classification of clouds are given in the Cloud Atlas published in 1957 by the Main Administration of the Hydrometeorological Service of the USSR, of which a summary table is given at the end of this book (see Appendix). This, the official atlas in the USSR, differs in certain details from the International Cloud Atlas published in 1956 by the World Meteorological Organization, but the principles of classification are the same.

Climatological processing of observations indicates a highly complex pattern of geographic distribution of cloudiness. The overall features are the following: the equatorial zone is characterized by considerable cloudiness throughout the year; in tropical zones (20–35° of latitude) cloudiness is minimum, and the principal minimum is observed over the continents, which contain the desert regions; cloudiness generally increases from the subtropics towards the poles to latitude 65–70 (or 75°); in regions close to the poles cloudiness drops slightly, remaining, however, considerable. Thus the zones of distribution of cloudiness coincide on the whole with the distribution of relative humidity.

Globally the mean annual cloudiness may be estimated as 5.4 points; this figure is somewhat lower over land (4.8 points) and higher over sea (5.8 points). The cloudiest places in the northern hemisphere are the northern parts of the Atlantic and Pacific Oceans and the Arctic Ocean; mean cloudiness here is over 8 points. The regions with the lowest cloudiness, as was mentioned above, are the subtropical deserts, which lie in a high-pressure zone. In the USSR the area with the greatest number of cloudless days is Central Asia, where cloudiness averages 1–2 points.

The geographic distribution of cloud forms has not been studied in sufficient detail. On the whole, however, one can state the following: clouds of vertical development (thick Cu and Cb) predominate in the equatorial zone owing to the prevalence of convection; in middle latitudes convective clouds predominate over land in summer and stratus forms in winter; in high latitudes stratiform clouds predominate almost exclusively.

The annual march of cloudiness is most clear-cut at middle latitudes but varies a great deal from region to region. Over the European USSR, for instance, maximum cloudiness and the maximum number of overcast days

usually occur in autumn and winter; in Siberia and Yakutia minimum cloudiness is recorded in winter and early spring due to the development of a pressure maximum at this time.

Diurnal periodicity is most clear-cut in the case of low and middle clouds. Stratus forms — St, Sc and partly As — are most often formed at night or towards morning and dissipate in daytime. Cumulus clouds usually develop in daytime and dissipate toward evening. Therefore over land in summer, when the prevailing cloud forms are convective, the maximum cloudiness is observed in daytime (afternoon).

Stratiform clouds are typical of the cold season, the maximum cloudiness being recorded in the norming hours. However, there may be two maxima of cloudiness during the day: one in the morning and the other in the afternoon. A daily march of this type is encountered most frequently in spring, when relative humidity is still high in the morning and radiation stratus is formed at this time, while in daytime convective clouds develop.

Over the ocean the daily march is reversed: convective clouds develop mostly at night when the water surface is warmer than the air.

The development and daily march of clouds are greatly influenced by local features. In mountain areas, for instance, particularly thick convective clouds develop in daytime in summer while mornings tend to be clear and cloudless; in coastal areas, on the other hand, cloudiness is lower in daytime.

§ 9. Altitude and vertical extension of clouds

The level of cloud base and the thickness (vertical extension) of clouds are highly variable in space and time even for the same cloud form.

The mean heights and thicknesses of the principal cloud forms have been determined for various geographic points from a great many measurements involving both aircraft flights and other aerological methods (pilot balloons, radiosonde, etc.). Values most characteristic of the middle latitudes are cited in the summary table of cloud classification (cf. Appendix II).

With regard to geographic differences in the height of clouds we note that the heights of practically all cloud forms rise from the high latitudes towards the equator. The heights of low and middle clouds display annual fluctuations: clouds are predominantly higher in summer than in winter.

In the warm season daily changes in the height of cloud base are also clear-cut: the heights are greater in daytime than at night. In individual cases the height of clouds may depart substantially from the mean, sometimes chaning within a short interval. Usually it is necessary to know the height of cloud base in each concrete case. For low clouds computational methods are used in the absence of direct measurements.

The simplest computational formulas were obtained under the assumption that cloud base coincides with the condensation level in moist air rising adiabatically from the surface layer. Ferrel's formula, cited earlier in Chapter 6, § 10, was obtained under these conditions:

$$Z_{con}=121\,(T_1-\tau_1).$$

Heights calculated from the above are usually in poor agreement with observational data. A number of attempts have been made to refine this

formula on the basis of results of aerological observations. Empirical formulas proposed for the cloud height H_{cl} have the same form as Ferrel's formula and differ from it only in the value of the coefficient

$$H_{cl} = a(T - \tau), \tag{27}$$

or else H_{cl} is expressed in terms of the relative humidity

$$H_{cl} = A(B - \log f) \tag{28}$$

(formula of D. L. Laikhtman et al.) and

$$H_{cl} = C(100 - f) \tag{29}$$

(formula of A. N. Ippolitov et al.).

As E. S. Selezneva has shown, these formulas will transform directly into each other if the difference $(T - \tau)$ is expressed in terms of f (cf. Chapter 6). Then for the adiabatic ascent of air one can write (for initial $T = 273°$)

$$H_{cl} = 123(T - \tau) = 3874(2 - \log f) \approx 17(100 - f). \tag{30}$$

However, adiabatic ascent of air is usually disrupted due to the heat and moisture exchange between the rising particle and the medium. Therefore the above formula can be used to a certain approximation only for convective clouds formed as a result of the ascent of considerable masses of air. For low stratiform clouds (St, Frst, Frnb, low Sc), which result from the turbulent upward transport of water vapor from the surface layer, the coefficients in formula (30) have other values. According to Selezneva they depend on the air temperature at the initial level and on the equilibrium temperature gradient. Thus for average conditions in the cold half of the year

$$H_{St} = 215(T - \tau).$$

A more rigorous solution of the problem taking turbulent transfer into account was given by M. E. Shvets. He obtained a formula of the form (29) and demonstrated that the coefficient C depends not only on γ_{eq} but also on γ_{av} as well as on the gradient of specific humidity β in the sub-cloud layer. In the absence of these data (γ_{av} and β) the coefficient C remains empirical.

It should be mentioned that cloud (like ground fog) becomes visible only when a certain amount of moisture has condensed (a definite concentration of liquid water has been reached). This requires lowering of the air temperature below the dew point by the additional amount $\Delta\tau$, which corresponds to ascent above the condensation level. When this factor is allowed for the formulas for H_{cl} take the form

$$H_{cl} = a[(T - \tau) + \Delta\tau],$$

or

$$H_{cl} = C[100 + \Delta f - f].$$

I. I. Chestnaya corrected Ippolitov's formula accordingly, writing it as follows

$$H_{cl} = 22(107 - f). \tag{31}$$

Selezneva gives an analogous relation for St cloud

$$H_{St} = 25(102 - f) = 5400(2.015 - \log f). \tag{32}$$

As to the vertical thickness of clouds, this varies over a very wide range for the different forms: from 50–100 m to several kilometers. The thinnest are the broken clouds (Frst and Frcu), which usually represent remnants of some cloud layer (St or Sc). Of the principal forms the thinnest are the high and middle wave clouds (Ac and Cc) and the thickest Ns and clouds of vertical development (thick Cu and Cb, with vertical extension of up to several kilometers). The vertical thickness of high clouds also varies fairly considerably. The most probable thickness of Ci and Cs (after A. M Barinov) is 1–2 km. Occasionally cirrus clouds virtually represent a zone of crystal precipitation bands; in such cases the clouds have considerable vertical extension, reaching 3–4 km or even more.

Considerable vertical extension (up to 7–10 km) typifies cloud systems developing at fronts. In most cases the tropopause marks the upper limit of cloud tops. However, with the development of high-altitude aviation it was found that in many instances cirrus clouds extended over both the tropopause and lower stratosphere.

Chapter 20

PRECIPITATION

Atmospheric precipitation, or simply precipitation, designates all the forms of water, liquid and solid, received from the atmosphere by the earth's surface.

The amount of precipitation is usually measured by the thickness (in millimeters or centimeters) of the layer of water which would form at the surface in the absence of percolation, runoff and evaporation.

An important characteristic of precipitation is its duration and intensity, i.e., the amount of precipitation (in millimeters of the water layer) per unit time (*minute or hour*).

In observations at meteorological stations a good deal of attenuation is also paid to the appearance of the falling particles; the first distinction to be drawn is between solid and liquid precipitation.

Water falling from clouds in the form of rain and snow accounts for the greater part of precipitation. A considerably smaller fraction is due to direct condensation and sublimation of water vapor on the surface from the adjoining layer of air; this is called surface deposition. In the present chapter we will consider the forms of precipitation and conditions leading to their formation.

§ 1. Surface deposition and conditions of formation

Surface deposition on the ground and on various objects is due to:

a) direct condensation or sublimation of water vapor,

b) deposition of water droplets contained in air flowing past.

Let us briefly consider the forms of surface deposition.

Dew ⏢ *and white frost* ⊔.* *Dew is a form of liquid deposition* consisting of small droplets of water: it forms at night, and sometimes in the evening, usually on open surfaces of soil, plant leaves, building roofs, etc.

Dew is formed by the condensation of water vapor when moist air comes into contact with a cooler surface. White frost, a white deposit of crystalline structure, is formed when the temperature of this surface is below 0°.

Weather conditions favoring the formation of dew or white frost are clear and comparatively calm nights when nighttime radiative chilling of soil and plants has a particularly marked effect on the daily temperature march. Mild winds favor the formation of abundant dew while strong winds accompanied by intensive mixing of lower chilled air layers and upper warm ones tend to level out their temperatures and thus prevent the condensation of dew

* Symbols are those used to designate forms of precipitation in meteorological observations.

Soil and objects with a low calorimetric conductivity are chilled to a greater degree at night and therefore dew and white frost are particularly heavy on loosened soil, building roofs, etc.

The amount of dew formed depends, of course, on the humidity of the air, increasing as the latter increases: however, at very high humidities effective emission decreases and the amount of heat released in condensation rises. From the quantitative standpoint the problem of dew formation is solved in a manner analogous to that of nighttime cooling and fog formation.

The theoretical considerations of M. E. Berlyand seem to indicate that the air layer involved in dew formation is 200–300 m thick (sometimes even thicker). This does not agree with the widespread view that dew is formed entirely from moisture contained in a thin (several meters) layer of lower air. The theory makes it possible to compute the amount of dew deposited over a certain time as a function of the parameters which determine the rate of dew deposition: atmospheric humidity, wind speed, soil humidity, etc.

On the average in middle latitudes dew yields a layer of about 0.1–0.3 mm per night and 10–50 mm over the year. Dew is heavier in autumn and spring since cooling below the dew point occurs more rarely in summer As the formation of heavy dew attenuates nighttime cooling of the air owing to the release of the heat of condensation, the presence of heavy dew reduces the probability of frost and radiation fog.

Liquid ⊼ and solid □ film. Liquid film is a deposit of water droplets — and sometimes a continuous film of water — formed chiefly on the windward side of objects cooled to temperatures considerably lower than the temperature of the surrounding air. In winter when the temperatures of surfaces are below 0° these deposits occur in the form of white semi-transparent crystals of ice; they are then called solid film

Unlike dew and white frost, liquid and solid film appear chiefly on vertical surfaces, building walls and stones. They are usually formed in overcast weather when a long period of cold is succeeded by sharp warming. They may be deposited at any time of day. The layer of solid film usually does not exceed a few millimeters in thickness,

Rime ⋎, V. Rime is a white, friable snowlike deposit formed on tree branches and shrubs, telegraph poles and wires and corners and projections of buildings. It often takes the form of fringes of fantastically-shaped ice needles. A light tremor suffices to make these fall. Sometimes rime freezes over and changes into a layer of loose, friable, opaque ice.

Rime differs from white frost in that it is observed at any time of the day, sometimes under cloudy skies but mostly in cold foggy weather or during severe frosts, when the air contains floating ice crystals. It differs from solid film in its delicate and friable structure.

Rime is formed either when minute supercooled droplets of fog come into contact with objects and freeze rapidly without spreading, in which case the rime layer will consist of faintly discernible ice grains (granular rime ⋎), or by direct sublimation of water vapor. The latter mechanism will produce a layer of minute ice crystals (crystalline rime V). Sublimation rime, unlike granular rime, is formed during stills or when winds are weak, and does not produce very thick deposits: the layer of sublimation rime usually does not exceed 1 cm in thickness whereas granular frost may reach 1 m in exceptional cases.

Glaze ∽. Glaze is a layer of dense ice, transparent or cloudy, appearing on horizontal and vertical surfaces predominantly on the windward side. On small objects, wires and tree branches glaze may be deposited on all sides.

Glaze is formed chiefly as a result of the freezing of large supercooled drops settling on the surface. It differs from solid film in its high intensity and in the fact that it forms a continuous, dense layer of ice. In the formation of glaze the supercooled drops have time to spread over the surface and fuse into a coating. The glass-like layer of ice which forms covers bridges, roads, building walls, wires, etc.

Glaze is usually observed in early and late winter during light frosts (down to −3, −5°). The thickness of glaze layers may reach several centimeters. Glaze and rime often cause great damage; icing fells trees, breaks telegraph poles, cuts wires and disrupts electric transmission lines.

§ 2. Characterization and classification of precipitation from clouds

Precipitation reaching the earth's surface from clouds can be subdivided firstly into solid, liquid and mixed.

Solid precipitation includes the following major forms:

Snow ✳, consisting of ice and snow crystals varying widely in shape but most frequently stellar, often combining to form large flakes. Large snow crystals reach 10 mm across while large flakes may be up to 8–10 cm in size.

Graupel (snow pellets) ⁂, consisting of fairly soft, opaque pellets of spherical shape, white or matt in color (diameter up to 2–5 mm).

Granular snow ⁂, smaller than graupel (diameter of grains is 1 mm).

Small hail △, consisting of solid particles (pellets) 2–5 mm in diameter: the surface is transparent with an opaque white kernel.

Hail ▲, pieces of ice of various shapes and sizes. Hailstones are as a rule considerably larger than ice pellets, the diameter being between 2–3 mm and several centimeters.

Ice pellets (sleet) ⍋, consisting of fine, solid, completely transparent spherical particles 1–3 mm in diameter.

Ice needles ↔, minute ice crystals without dendritic structure. On frosty days they are seen as flashes sparkling in the sun.

Liquid precipitation includes the following forms.

Rain •, precipitation in the form of water droplets of varying size (from 0.5 to 7 mm in diameter).

Drizzle ❟, consisting of raindrops so minute (diameter < 0.5 mm) that their fall is nearly imperceptible and they seem suspended in air.

Finally, precipitation of mixed character is not infrequent; this consists of a mixture of solid precipitation and rain falling at the same time (e. g., ice pellets and rain). Also frequent is the so-called wet snow ⁎ — precipitation consisting of melting snow or a mixture of snow and rain.

Depending on physical conditions of formation, duration and intensity, precipitation is described as·

1) steady precipitation, falling from frontal clouds (Ns and As) It is of medium intensity but is prolonged, covers a considerable area at a time and consists of medium-sized raindrops or snow;

2) showers, usually from Cb. They are frequently characterized by high intensity They begin suddenly and end quickly but frequently recur, usually covering a small area. Raindrops in showers are usually larger than in steady rain. In the solid form showers are composed of large snowflakes, graupel and sometimes hail. Their fall is frequently accompanied by electrical storms and squalls.

3) drizzle, from St or Sc. This consists of the smallest raindrops and finest snow crystals or ice needles. The intensity is very low. This variety forms chiefly in uniform, stable air masses.

§ 3. Size and fall velocity of particles of precipitation

Observations show that the radius of raindrops can vary between very broad limits, roughly from 0.1 to 3.5 mm The radius of falling drops does not tend to be greater than 3.5 mm since for larger sizes the fall velocity increases, leading to deformation and disintegration of the drops. For the indicated drop sizes the mass varies roughly between $5 \cdot 10^{-6}$ and $1.5 \cdot 10^{-1}$ g.

As to the linear dimensions of particles of solid precipitation, these are extremely variable, especially in the case of snow crystals and flakes, and also hailstones. Thus large flakes may reach 15–20 cm in diameter and spherical hailstones 10–15 cm.

Owing to their substantial departures from sphericity, it is usual to characterize the size of particles of solid precipitation (e. g., snow crystals) by their masses. Observations show that small crystals and flakes have masses of the order of 10^{-6} g, i. e., the same as small raindrops. Large flakes may reach 0.4–0.6 g and large hailstones several grams; in exceptional cases the weight of a hailstone may amount to several hundred grams.

The size distribution of raindrops and mass distribution of snow crystals are described in the same way as the cloud-droplet distribution, by means of various empirical formulas: an example is a formula analogous to the Khrgian-Mazin formula

$$f(r) = Br\alpha e^{-\beta r}, \tag{1}$$

where B, β and α are empirical coefficients.

Like cloud particles, particles of precipitation have different fall velocities depending on their mass and form. If they fall in an updraft their fall velocity with respect to the earth s surface will be equal to the difference between their fall velocity in immobile air v and the vertical velocity of the ascending current w.

The question of the fall velocity v of spherical water drops of small radius r, for which Stokes' formula is applicable, was considered in Chapter 2, § 8. The fall velocity of such drops is given approximately by the formula

$$v = 1.26 \cdot 10^6 r \text{ cm sec.} \tag{2}$$

But, as experiments show, this formula is valid only for small cloud droplets (radius smaller than 50μ). For larger droplets and particles of precipitation it does not hold well. It has been established that the latter fall more slowly. For $r = 100\mu$, for instance, the velocity v according to

Stokes' formula is 126 cm/sec as against only 78 cm/sec according to experimental data.

Figure 120 shows two curves: 1–based on calculations, 2–on experimental data. Comparing the two one sees that a marked divergence is already present for r of the order of 50μ. This divergence is due to the fact

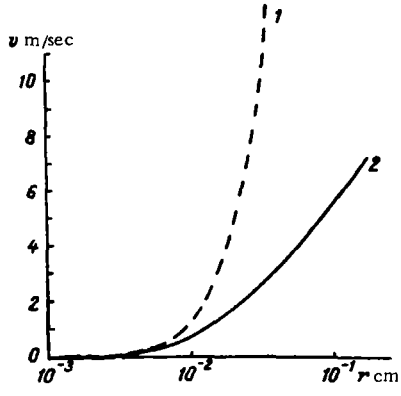


FIGURE 120. Fall velocities of droplets from Stokes' formula (1) and from experimental data (2)

that the resistance of the air to the falling drop is not accounted for sufficiently accurately in Stokes' formula. It can be made more exact by taking a more correct expression for the air resistance

$$F = C_D \pi r^2 \frac{\rho_a v^2}{2}. \tag{3}$$

Here C_D is the drag coefficient, which depends on Reynolds' number; this was found from experiments in wind tunnels.

Taking formula (2) into account, K. S. Shifrin obtained the following expression for v

$$v = \frac{12}{2\rho_a r} \Phi(x), \tag{4}$$

where $\Phi(x)$ is a function of the parameter

$$x = \alpha r^3 = \frac{4}{9} \frac{\rho_a \rho_d g}{\eta} r^3. \tag{5}$$

Here ρ_a and ρ_d are the density of the air and the falling spherical particle, g is the acceleration of gravity and η is the viscosity.

Formula (4) goes over into Stokes' formula for $x \ll 1$, while for large x one obtains the formula proposed by L. Krastanow

$$v = a\sqrt{r}, \tag{6}$$

where $a = (1.5-2) \cdot 10^3\ \text{cm}^{1/2}/\text{sec}$.

This formula is valid for $r > 500\mu$.

Table 64 gives experimental values of the stable fall velocity of water drops in air for $p = 1000$ mb and $t = 20°$.

From the above discussions it is evident that small cloud droplets with radii smaller than 20μ rise rather than fall when the velocity of the ascending air flow $w > 5$ cm/sec. In order for droplets to fall it is necessary that w be smaller than v. Sometimes updrafts of the order of several m/sec are encountered in clouds; in such cases even the largest drops ($r = 3$–3.5 mm) are unable to fall, and even larger ones ($r > 3.5$ mm) are deformed and disintegrate for $w \approx 9$ m/sec.

TABLE 64

Fall velocities of cloud droplets and raindrops

r mm	0.02	0.05	0.1	0.2	0.5	1.0	1.5	2.0	2.5	3.0
v (cm/sec)	5	27	72	162	403	649	806	883	909	918

The question of the fall velocities of solid particles of precipitation is far more complicated. For spherical particles the fall velocity can be computed in the same way as that of water droplets, making due allowance for the fact that their density $\rho < 1$. Owing to the fact that falling ice crystals and snowflakes continuously change position in order to encounter maximum air resistance, it becomes difficult to compute their fall velocities. Observations show that platelike or stellar crystals with the same mass as droplets fall more slowly than droplets. Thus ice needles ($l \approx 1$ mm) have a velocity of about 0.5 m/sec and rimed crystals ($r \approx 2$ mm) of about 1 m/sec. Particles of graupel and especially hail with masses considerably larger than that of the largest raindrops fall very rapidly. For example, graupel with $r > 1$ mm has a fall velocity of about 2 m/sec; hailstones with $r = 0.6$ cm fall at a velocity of about 10 m/sec and those with $r = 2.5$ cm at about 25 m/sec; larger hailstones have even higher fall velocities.

§ 4. Evaporation of raindrops

Like other particles of precipitation, raindrops emerging from clouds enter air layers in which the relative humidity is generally less than 100%. Evaporation takes place from the surface of the falling particles and sometimes leads to their complete evaporation; drops which fail to reach the earth's surface form the so-called precipitation bands (virga).

The process of evaporation plays a significant role in the delivery of rain to the earth's surface, and is therefore worth examining in detail.

Let us assume that a water drop of radius r_d occurs in a medium in which the vapor concentration at a certain distance $r \gg r_d$ from the center of the drop is a_{air}. If the vapor concentration at the surface of the drop $a_d > a_{air}$, then the following flux of water vapor will spread out from its entire surface ($4\pi r_d^2$) into the surrounding space:

$$W = -4\pi r_d^2 D \frac{da}{dr}, \tag{7}$$

where D is the coefficient of diffusion of water vapor and $\frac{da}{dr}$ is the gradient of the water vapor concentration.

This total flux of vapor will pass through every spherical domain surrounding the drop and therefore one can take $W(r) = \text{const}$.

Let us separate the variables and integrate expression (7) from r_d to r and from a_d to a_{air} Obviously,

$$a_d - a_{air} = \frac{W}{4\pi D}\left(\frac{1}{r_d} - \frac{1}{r}\right). \tag{8}$$

Recalling that $r \gg r_d$, we have, approximately,

$$W = 4\pi D(a_d - a_{air})r_d. \tag{9}$$

Since the droplet mass is $m = \frac{4}{3}\pi r_d^3 \rho_d$, where ρ_d is its density,

$$\frac{dm}{dt} = -W = 4\pi r_d^2 \rho_d \frac{dr_d}{dt},$$

whence the variation in the droplet radius with time, bearing (9) in mind, will be

$$\frac{dr_d}{dt} = -\frac{W}{4\pi r_d^2 \rho_d} = -\frac{D}{\rho_d}\frac{(a_d - a_{air})}{r_d}. \tag{10}$$

From this expression it is evident that the rate of change of droplet size by evaporation is inversely proportional to its radius, i.e., the larger the drop the slower the rate at which its radius decreases for the same difference $(a_d - a_{air})$.

Let us determine the time required for the initial droplet radius r_1 to decrease to r_2. We will confine ourselves to a simplified derivation, assuming that the buoyancy of saturated vapor over the drop does not depend on the curvature of its surface (i.e., we will take $a_d - a_{air} = \text{constant}$). Then from (10), separating the variables and integrating over the time from zero to t and over r from r_1 to r_2, we find that

$$r_1^2 - r_2^2 = \frac{2D}{\rho_d}(a_d - a_{air})t, \tag{11}$$

whence

$$t = \frac{\rho_d}{2D}\frac{r_1^2 - r_2^2}{(a_d - a_{air})}. \tag{12}$$

Passing from the volume concentration of water vapor a g/cm^3 to its buoyancy e mb (to do this we employ the relation $a = \frac{e}{R_{vap}T}$), we obtain

$$t_{r_1 \to r_2} = c_{ev}\frac{T}{T_0}\frac{r_1^2 - r_2^2}{E_d - e_{air}}. \tag{13}$$

For $\rho_d = 1$ g/cm^3, $R_{vap} = 4.60 \cdot 10^6$ erg/g $\cdot$ deg, $D = 0.22$ cm^2/sec and $T_0 = 273°$,

$$c_{ev} = \frac{\rho_d R_{vap} T_0}{2D} = 2.85 \cdot 10^6 \text{ mb} \cdot \text{sec/cm}^2 .$$

According to the experimental data of N. Frössling, E. G. Zak, B. V. Kiryukhin, N. P. Tverskaya et al., $c_{ev} = 3 \cdot 10^6$, which agrees with the value given above.

Another important point must be mentioned in connection with the question of evaporation. The heat lost in evaporation causes a lowering of the temperature of the drop. In turn the temperature difference between the

drop and surrounding air creates a flow of heat toward the surface of the drop.

In the stationary state the heat expended in evaporating the drop WL is compensated by the influx of heat Q from the air, i. e.,

$$WL = -Q = +\lambda \frac{\partial T}{\partial r},$$

where L is the heat of evaporation, λ is the calorimetric conductivity and $\frac{\partial T}{\partial r}$ is the temperature gradient between the drop and the air.

Using the same method as when determining the flux of water vapor, we obtain

$$4\pi DL(a_d - a_{air})r = -4\pi\lambda(T_d - T_{air})r, \tag{14}$$

whence

$$T_d = T_{air} - \frac{DL}{\lambda}(a_d - a_{air}),$$

where a_d should correspond to the temperature of the drop.

Approximately one can say that the temperature of the drop is equal to the wet-bulb temperature of a psychrometer. When the drop temperature is taken into account (after Mason) expression (10) assumes the more complicated form

$$r_1 \frac{dr_d}{dt} = \frac{a_d - a_{air}}{\frac{AL^2}{\lambda T^2} + \frac{T}{a_d D}}. \tag{15}$$

So far we have assumed the drop to be immobile with respect to the surrounding air. In reality when drops fall in the atmosphere currents of air flow around them. This makes it necessary to introduce into (15) the correction factor $f = 1 + F$, where F is a complex function of the size of the drop and its fall velocity. Then formula (13) becomes

$$t_{r_1 \to r_2} = c_{ev} \frac{T}{T_0} \frac{r_1^2 - r_2^2}{[E_d - e_{air}](1 + F)}. \tag{16}$$

It was found that $F = k\sqrt{\mathrm{Re}}$, where k is a certain constant ≈ 0.22 and $\mathrm{Re} = \frac{2v\rho_d r}{\eta}$ is Reynolds' number.

The theoretical derivation of the expression for the factor f, due to L. E. Leibenzon, and its experimental verification, due to N. P Tverskaya et al., show that for small droplets falling very slowly (radius less than 100μ) this correction may in practice be disregarded; for larger (rain) drops, on the other hand, it plays an important role and for (say) $r = 1000\mu = 10^{-1}$ cm the value of F amounts to about 5, which makes evaporation several times faster.

From the above formulas it is possible to find the time required for the complete evaporation of a drop (zero radius)

For illustration, Table 65 gives the complete evaporation times t for droplets of various sizes with $= 90\%$ and $t = 5°$, as calculated from formula (16).

One can see that the time required for the evaporation of minute droplets is very small and that they disappear in a few seconds. Larger drops

($r > 100\mu$) evaporate more slowly and are therefore capable of travelling over considerable distances and reaching the earth's surface.

TABLE 65

Complete evaporation times of drops

$r_1\,\mu$	1	10	100	1000
t	0.05 sec	5.3 sec	6.1 min	2.9 hrs

Let us evaluate the distance travelled by a falling droplet (initial radius r_1) before its complete evaporation. If the fall velocity of the drop with respect to air is v, then in the time dt it will travel over the path $dz = vdt$.

From expression (10), which includes the correction for air currents, we have

$$dt = -\frac{\rho_d}{D}\frac{rdr}{[a_d - a_{air}][1+F]}. \qquad (17)$$

Introducing the value of dt into the preceding expression for the path, we obtain

$$dz = -\frac{\rho_d}{D}\frac{rdr}{[a_d - a_{air}][1+F]}v \qquad (18)$$

(the minus appears because the z-axis and velocity v have different directions [sic]).

However, the task of determining the path travelled by the drop before complete evaporation from this expression is complicated by the fact that,

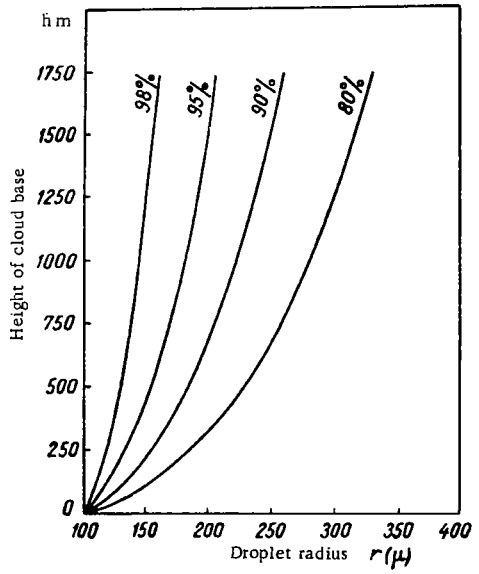


FIGURE 121. Evaporation of droplets falling from cloud base for different values of the relative humidity (after Mason)

apart from the temperature and humidity, the fall velocity v as well as the factor F vary along the path of the drop. For an approximate estimate of z

mean values of the quantities appearing in the formulas are employed. Such estimates indicate that even at a comparatively high humidity (90%), in the absence of vertical movements of air only fairly large drops ($r > 200\mu$) will be able to travel a distance of the order of one kilometer.

The curves given in Figure 121 indicate the heights from which droplets must fall in order to evaporate to the radius $r = 100\mu$; the calculations were carried out for various mean values of the relative humidity and temperature.

If ascending currents of velocity w are present in the atmosphere, the fall velocity will slow down and for $v < w$ the droplets will move upwards. Under these conditions the path travelled by a drop relatively to the earth's surface will be

$$dz = (v - w)\,dt. \tag{19}$$

Correspondingly equation (19) will take the form

$$dz = -\frac{\rho_d}{D(a_d - a_{air})}\frac{r\,dr}{}(v - w). \tag{20}$$

Calculations according to this formula show that, for $w = 1$ m/sec, $T = 288°K$ ($t = 5°$) and $f = 90\%$, a drop of radius $r = 200\mu$ will evaporate completely on a path $z_1 = 500$ m. For $w = 1$ m/sec drops with $r > 500\mu$ will be able to travel a distance greater than 1 km.

§ 5. Growth of droplets in clouds

For the formation of precipitation it is necessary that the cloud particles grow in size until they are able to fall from the cloud. Further, it is necessary that their fall velocity be greater than the velocity of the updrafts and that they can travel the entire distance from cloud to ground without evaporating.

The principal mechanisms by which particles increase in size within clouds are coalescence (accretion) and condensation or sublimation of water vapor. Consequently,

$$\frac{dm}{dt} = \left(\frac{dm}{dt}\right)_{cond} + \left(\frac{dm}{dt}\right)_{acc}. \tag{21}$$

Growth of droplets by condensation. The main factor at the initial stages of growth of cloud particles is condensation and sublimation. For these processes to develop it is necessary that the buoyancy of vapor in the air be greater than the buoyancy of saturated vapor over the surface of the droplets, i. e., that there be a certain supersaturation of water vapor with respect to the droplets in the cloud.

In exploring the role of condensation we will make use of the results of the previous section, bearing in mind, however, that we are now dealing with the growth of droplets due to condensation when $a_{air} > a_d$. In this case the rate of increase of mass of a droplet of radius r_d is

$$\left(\frac{dm}{dt}\right)_{cond} = 4\pi D(a_{air} - a_d)r_d \tag{22}$$

and

$$\left(\frac{dr}{dt}\right)_{\text{cond}} = \frac{D}{\rho_d}\frac{(a_{\text{air}} - a_d)}{r_d}. \qquad (22')$$

Passing over to the buoyancy of water vapor $a_d = \frac{E_d}{R_{\text{vap}}T}$, we obtain

$$\left(\frac{dr}{dt}\right)_{\text{cond}} = c_{\text{cond}}\frac{(e - E_d)}{r}\frac{T_0}{T}, \qquad (23)$$

where $c_{\text{cond}} = \frac{D}{\rho_d R_{\text{vap}} T_0}$. Numerically $c_{\text{cond}} = 1.8 \cdot 10^{-7}$ cm^2/mb · sec.

To illustrate the calculation of the rate of condensational growth of droplets, we will take T_d equal to the air temperature $T_{\text{air}} = 273°$K and the relative humidity equal to 101 % in one case and 100.1 % in the other, which gives, respectively, 0.04 and 0.004 mb for $e - E_d$.

The calculated results are given in Table 66 (times in which a droplet with initial radius $r_0 = 1\,\mu$ will increase its radius by a factor of 2, 10, 100 and 200).

TABLE 66

$\frac{r}{r_0}$	2	10	100	200
f = 101 %	2.1 sec	75 sec	2 hrs	8 hrs
f = 100.1 %	21 sec	12.5 min	20 hrs	—

As one can see, small droplets grow fairly rapidly at first: in a few seconds their radius increases 2–3 times. As the droplet size increases their growth slows down, and at the limited supersaturations possible in clouds the formation of raindrops ($r > 100\,\mu$) requires several hours. Since the larger drops grow very slowly with time, condensation in clouds should result in the appearance of droplets of more uniform size.

The above formulas make it possible to compute the rate of growth of droplets and their size at any instant if the supersaturation of water vapor in the cloud is known. So far, however, it has not been possible to measure this quantity with the required accuracy and it is usually estimated from theoretical considerations.

Growth of droplets by accretion. This is due to collisions between droplets, which can result in their fusion (coalescence). Examination of the question of coalescence in clouds shows that particles may approach and collide with each other for various reasons. In the formation of particles of precipitation, however, the main factor is collision between droplets having different sizes and therefore different fall velocities. This is the so-called gravitational coalescence.

Coalescence owing to the Brownian motion of particles and due to turbulent motions, as calculations show, is important only for very small particles (radius up to 1–2 μ) in large numbers. Thus to double the particle radius by Brownian motion requires about 30 min for an initial radius $r_{\text{in}} = 1\,\mu$ and a liquid-water concentration $a_{\text{liq}} = 1$ g/m^3, whereas about 10 hrs are needed to increase the radius to $r = 7\,\mu$ for $r_{\text{in}} = 6\,\mu$.

We note that a significant growth of a droplet by accretion involves the coalescence of many droplets. Indeed, the ratio of the volumes of two droplets of different size is $\frac{V_2}{V_1} = \left(\frac{r_2}{r_1}\right)^3$: therefore, if (say) all the droplets have

the radius $r_1 = 5\mu$, the formation of a raindrop with $r_2 = 500\mu$ as a result of the coalescence of such droplets will require about a million small cloud droplets.

Since for the formation of precipitation gravitational coalescence is paramount and all other forms of coalescence may be disregarded, we will consider only the former in detail.

A very important problem in the theory of gravitational coalescence is that of the aerodynamic conditions of droplet collision. Let there be a large drop of radius R lying in a stream of air containing smaller droplets (Figure 122). At great distances from the large drop the trajectories of the air particles, as well as of the small droplets, will consist of parallel lines. But at a certain distance from the large drop the trajectories of the air particles will begin to bend around the large drop. The smallest droplets, which in practice move with the air particles, will also flow around the large drop and will not collide with it. Large droplets with considerable inertia, however, will be deflected from the trajectory of the air particles and may collide with the large drop. Many works have been devoted to the theoretical consideration of the question of aerodynamic conditions of collision of spherical particles (Langmuir, N. S. Shishkin, L. M. Levin, et al.).

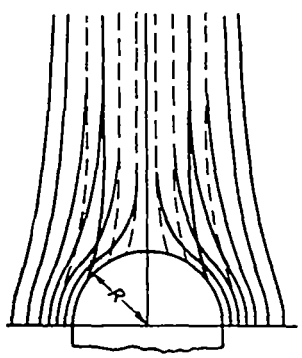


FIGURE 122. Streamlines and inertial trajectories of droplets

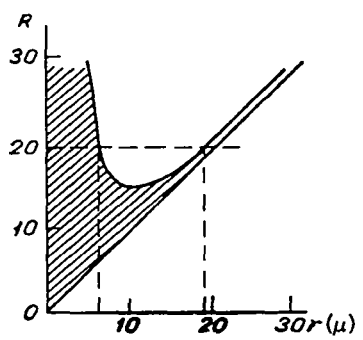


FIGURE 123. Region of accretion between drops of different sizes

The ratio of the number of small droplets colliding with the large drop to the number of all droplets which would have collided with the large drop if they had been moving rectilinearly all the time is called the collision efficiency ε'. It depends on the radius R of the large drop and radius r of the small droplets and, according to the calculations of Langmuir, is given by

$$\varepsilon'(R, r) = \left(1 + \frac{0.75 \ln 2k}{k - 1.214}\right)^{-2}, \qquad (24)$$

where $k = \frac{\lambda}{R}$ is the so-called coefficient of inertia and λ the inertial path of the droplets.

If the initial velocity of a droplet of radius r is v_0 at the point at which it begins to deviate due to inertia from the air-particle trajectory, its velocity will decrease upon further motion owing to the air resistance. If

we take the resistance of the medium to be, according to Stokes,

$$F = -6\pi\eta r v,$$

then the reduction in velocity will take place as follows

$$v = v_0 e^{-\frac{t}{\tau}},$$

where $\tau = \frac{2}{9}\frac{\rho}{\eta} r^2$ is the relaxation time, i.e., the time in which the velocity of the droplet decreases e times. Then the path travelled by the droplet due to inertia will obviously be

$$\lambda = \int_0^\infty v_0 e^{-\frac{t}{\tau}} dt = v_0 \tau. \tag{25}$$

In calculating the collision efficiency Langmuir regarded the small droplets (radius r) as immobile and the large drop (radius R) as falling at a uniform velocity. However, as Shishkin pointed out, when calculating ε' one must account for the relative velocity of the large and small drops.

Calculations show that the small droplets, in practice; do not experience mutual collisions.

For an approximate evaluation of the collision efficiency one can use the simpler expression suggested by Shishkin:

$$\varepsilon' = \left(\frac{k - 1.214}{k}\right)^2, \tag{26}$$

where $k = \frac{\lambda}{R}$, as before

Corresponding calculations show that gravitational coalescence begins only when some of the drops reach the size $R \geqslant 14.5\mu$. For instance, droplets of radius $r = 4\mu$ cannot collide with drops of radius $R \leqslant 48\mu$, and droplets of radius $r = 6\mu$ cannot collide with drops of radius $R \leqslant 23\mu$. Figure 123 shows the limits of variation of the radius r of those droplets which can coalesce with drops of radius R.

In using the above relations to calculate the change in drop size due to coalescence it should be remembered that far from every collision necessarily leads to coalescence. The explanation for this lies in the formation between the colliding particles of a thin layer of air in which excess pressure (of the order of 1 mb) counteracts coalescence.

Experimental studies of this complex question by B. V. Deryagin, P. S. Prokhorov and others show that the coalescence efficiency depends to a very large extent on the relative humidity of the surrounding air, and also on an entire series of secondary factors which include: relative velocity at the instant of collision, nature of the collision and presence of electric charges on the drops. At humidities of 100% practically all collisions lead to coalescence.

The proportion of droplets coalesced with the larger drop out of the number colliding with it is called the coalescence efficiency ε''. The product of the two efficiencies — collision and coalescence — is known as the collection efficiency $\varepsilon^* = \varepsilon'\varepsilon''$.

Since, however, at a humidity of 100% (in cloud and fog) all collisions usually lead to coalescence, in these ε'' is close to unity and $\varepsilon^* = \varepsilon'$, i.e., the collection efficiency can be regarded as equal to the collision efficiency.

Let us now consider the question of the growth of a drop of radius R as

a result of collisions with droplets of smaller radius r. While falling through the cloud this drop will move over the path dz relative to a droplet of radius r in the time dt:

$$dz = (v_R - v_r)\,dt. \tag{27}$$

In this process the large drop captures smaller droplets in a volume of cross section $S = \pi (R+r)^2$.

If the liquid-water content of the cloud is a_{liq} and the collection efficiency ε^*, then the increase in mass of the growing drop is

$$\left(\frac{dm}{dt}\right)_{acc} = \varepsilon^* \pi (R+r)^2 (v_R - v_r)\, a_{liq} = 4\pi\rho_d R^2 \frac{dR}{dt},$$

whence

$$\frac{dR}{dt} = \varepsilon^* \frac{(R+r)^2}{R^2} (v_R - v_r) \frac{a_{liq}}{4\rho_d}. \tag{28}$$

Recalling that

$$\varepsilon^* = f(R, r) \text{ and } a_{liq} = \sum_i m_i N_i = \frac{4}{3}\pi\rho_d N \int_0^\infty r^3 n(r)\,dr,$$

it would be more exact to write (28) as

$$\frac{dR}{dt} = \frac{1}{R^2} \frac{a_{liq}}{4\rho_d} \frac{\int_0^R \varepsilon^* (R+r)^2 (v_R - v_r) r^3 n(r)\,dr}{\int_0^\infty r^3 n(r)\,dr}. \tag{29}$$

Using approximate expressions, we obtain

$$\frac{dm}{dt} = \left(\frac{dm}{dt}\right)_{cond} + \left(\frac{dm}{dt}\right)_{acc} = 4\pi D (a_{air} - a_d) r + \\ + \varepsilon^* \pi (R+r)^2 (v_R - v_r)\, a_{liq} \tag{30}$$

We considered the process of drop growth without allowing for updrafts in the cloud. If an updraft is present the path travelled by the drop relative

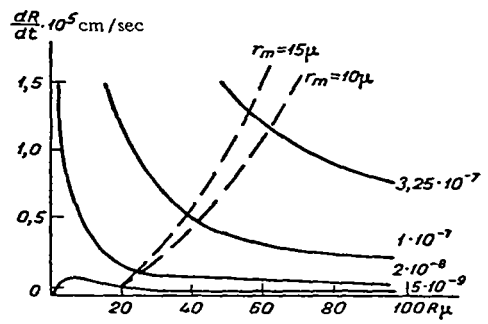


FIGURE 124. Rates of cloud-droplet growth by accretion and condensation for $a_{liq} = 1$ g/m^3 (after Shishkin)

to the ground will be

$$dz = (w - v)\,dt,$$

where w is the average speed of the updraft.

Then

$$dm = \left[\frac{4\pi D(a_{air} - a_d) r}{w - v_R} + \frac{\varepsilon^* \pi (R+r)^2 (v_R - v_r) a_{liq}}{w - v_R}\right] dz. \qquad (31)$$

The dashed curves in Figure 124 give the growth rates of droplets in a cloud with a mean liquid-water content $a_{liq} = 1\ g/m^3$ for r_m equal to 10 and 15μ (r_m is the radius of those droplets making the largest contribution to the liquid-water content of the cloud). In these calculations, which were performed by integration, the deviation of the fall velocity of the large droplet from the Stokes velocity was taken into account. The solid curves in the same figure give the rates of droplet growth due to condensation for various values of the supersaturation

$$\Delta e = \frac{a_d - a_{sat}}{\rho_{air}}.$$

According to the figure, as the drop radius increases the rate of condensational growth decreases while that of growth by accretion increases. The droplet radius at which growth by accretion predominates over condensational growth depends on the supersaturation, increasing with the latter. Thus, for supersaturations of the order of $\Delta e = 5 \cdot 10^{-9}$ (observed for updraft speeds of about 1 m/sec) the rate of growth by accretion surpasses condensational growth for $R = 20\mu$, while at high supersaturations this happens only for $R \approx 50$–70μ. Further, one can see that droplets with radii considerably greater than r_m will grow more rapidly by accretion than the smaller droplets.

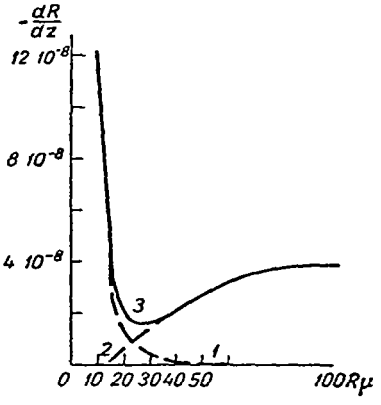


FIGURE 125. Growth rates of droplet falling through cloud

1–rate of growth by condensation; 2–rate of growth by accretion; 3–total growth rate.

Figure 125 gives the calculated growth rate of a droplet falling in a cloud in which the liquid-water content $a_{liq} = 0.5\ g/m^3$, the speed of the updraft $w = 0.5$ cm/sec, the droplet concentration $n = 5.3 \cdot 10^2\ cm^{-3}$ and the weighted mean of the droplet radius $\bar{r} = 4.1\mu$.

The expressions given above can be used to compute the growth rates of cloud droplets in various cases. This is illustrated in Figure 126 which shows the vertical change in the radius of droplets falling from a cloud at the beginning of a rain (for a cloud with a steady updraft). In calculating this it was assumed that the base of the cloud from which height was reckoned was at the 1000 m level, the temperature there being 6°. Growth rates were computed for a droplet having a radius of 10μ at cloud base for w varying from 10 to 100 cm/sec.

As one can see, the droplet first rises (carried by the air current), its size increasing in the process to a certain value $r = r_{peak}$ at the so-called peak of the trajectory, at which $v_R = w$. Next it begins to fall ($v_R > w$) and, while continuing to grow, falls out of the cloud; its final radius $r = r_{fin}$.

At the initial stage, along the upward path, the main factor is condensational growth; subsequently, starting from a certain level, growth takes

place chiefly due to accretion, especially when the droplet falls (descending branch of trajectory).

From the graph one can see how strongly the final radius r_{fin} of a raindrop emerging from a cloud depends on the velocity w of the ascending currents. Thus in the given case $r_{fin} = 0.33$ mm for $w = 10$ cm/sec while for $w = 100$ cm/sec r_{fin} already reaches values of 1.6 mm. It is also evident that for the droplet to fall from the cloud the cloud must have a definite thickness depending on the value of w. In the case under consideration, for instance, for $w = 10$ cm/sec the peak of the trajectory must lie at a height of about 1 km above cloud base, whereas for $w = 100$ cm/sec droplets must rise to a height of about 2.4 km.

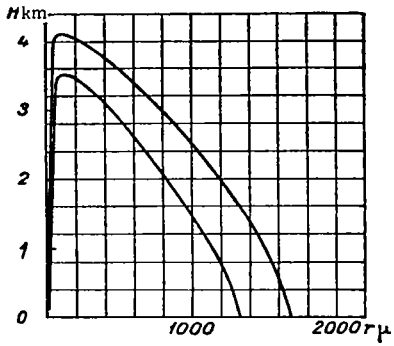


FIGURE 126. Growth of earliest raindrops

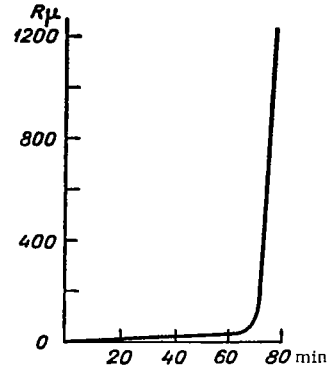


FIGURE 127. Growth of droplet as a function of time, in cloud with updraft of 70 cm/sec

As a second example Figure 127 presents results of calculations of droplet growth as a function of time for a cloud with an updraft $w = 70$ cm/sec. At first the droplet grows very slowly by condensation; very rapid growth by accretion begins only after radii of the order of 50–70 μ have been reached. Thus while the growth of the droplet from 10 to 50 μ requires about 30 min, its growth from 100 to 1200 μ requires only 6 min. Naturally, for smaller values of w the process will be slower.

The above data show that accretion is the basic mechanism of growth of droplets. From equation (31), considering only the accretion for large droplets with $R \gg r$ and $dm = 4\pi\rho_d R^2 dR$, we obtain the following on the descending branch of the trajectory

$$\frac{dR}{dz} = \varepsilon^* \frac{a_{liq}}{4\rho_d} \frac{v_R}{w - v_R}.$$

If $v_R \gg w$ then, integrating the above from $z = 0$ to $z = z_{peak}$ and from $R = r_{peak}$ to $R = r_{fin}$, we find

$$r_{fin} = r_{peak} + \varepsilon^* \frac{a_{liq}}{4\rho_d} \Delta z, \qquad (32)$$

where r_{fin} and r_{peak} are the radii of the droplet at the end and at the peak of the trajectory.

For large drops with $R > 1000\,\mu$ ($\varepsilon^* = 1$), assuming $a_{liq} = 1\,g/m^3$, the rate of droplet growth is approximately

$$\frac{\Delta r}{\Delta z} = \varepsilon^* \frac{a_{liq}}{4\rho_d} \approx 250\ \mu/km. \qquad (33)$$

§ 6. Precipitation from droplet clouds

Let us consider the formation of precipitation from various forms of droplet cloud.

In view of the importance of the cloud thickness and speed of the updrafts for the growth of raindrops, clouds with feeble and strong vertical currents will be considered separately.

Typical examples of liquid clouds in which the speed of updrafts is low and which yield occasional light precipitation of the drizzle type are stratus and stratocumulus; the thickness of these clouds is, as a rule, limited (less than 1 km).

In this case the origin of the precipitation can be explained as follows. Usually such clouds consist of fine droplets, with a predominant drpolet radius of less than 10μ. If they exist for a long time the process of condensation, though acting slowly, may eventually lead to the formation of a certain small number of larger droplets ($r \approx 20-30\mu$) which for small w will move downwards and grow by accretion while traversing the cloud layer. In clouds of limited thickness (about 1 km) these falling droplets are unable to grow to a large size and at cloud base they assume dimensions characteristic of drizzle ($r = 100-200\mu$); precipitation reaching the ground from such clouds is usually of low intensity.

In droplet clouds formed in the prolonged ascent of air at relatively high speeds, such as (in summer) altostratus and nimbostratus, conditions are more favorable for rain formation. The great vertical extension ensures the formation and precipitation of larger droplets having $r \approx 0.5-0.7$ mm.

If there is a constant delivery of water vapor into the cloud to compensate for the loss of moisture due to rainfall, precipitation from such clouds, as observations show, may continue for many hours and even twenty-four hours or more.

Let us now consider convective droplet clouds. These include the cumulus clouds (Cu, Cu cong). All these clouds are characterized by a high vertical speed of air movement, up to 6–8 m/sec and sometimes even more. The great updrafts, which lead to sharp chilling of the air, may create considerable supersaturations, ensuring the appearance of a large number of fine droplets. These currents lift not only the small droplets but fairly large drops as well. For vertical speeds of air movement of 5–8 m/sec only drops with $r \geq 1-2$ mm will fall out of the cloud. However, drops will grow to this size only if the liquid cloud is sufficiently thick — over 3 km. In middle latitudes, therefore, cumulus clouds having a vertical extention of 2–3 km do not usually produce precipitation.

When the summit of the convective clouds reaches high altitudes (8–10 km) while the base lies at a height of 2–3 km (a frequent situation in the tropics), conditions are particularly favorable for the formation of very large drops and consequently for the release of a heavy rain or shower. Droplets in such

clouds can grow to a limiting size ($r \approx 3.5$ mm); when this is reached in updrafts $w \approx 9$ m/sec the droplets disintegrate. The larger fragments may fall as raindrops while the smaller ones are lifted upwards. This process (according to Langmuir) gives rise to a special sort of "chain reaction". The small fragments, which are considerably larger than the mean cloud-droplet size, may again reach the critical radius and break up into raindrops and particles larger than cloud droplets.

The conditions required to maintain such a chain reaction are fairly complicated. A great cloud thickness (5–6 km) and liquid-water content of not less than 2–3 g/m^3 are necessary. These conditions are frequently encountered in droplet clouds of the tropical zones but are rare in middle latitudes.

§ 7. Precipitation from ice and mixed clouds

Compared with droplet clouds, conditions of precipitation in ice and mixed clouds are far more complicated, though at the same time more favorable. All observations indicate that at high and middle latitudes practically all intense precipitation emerges from mixed clouds in which ice crystals are present (if only in the upper parts).

In the formation of precipitation from ice-crystal and mixed clouds the growth of crystals by sublimation is more significant than the condensational growth of droplets. The reason is that, in such clouds, at temperatures ranging down to −40°, the vapor buoyancy (concentration) usually observed corresponds to saturation over water rather than over ice. This being the case there is a considerable supersaturation (several tens of percents) with respect to ice and ice particles in the cloud grow much more rapidly than the water droplets. This growth is most intensive at temperatures of about −12.5° at which, as we saw, the difference in the vapor buoyancies over ice and water is maximum.

Quantitatively the question of the sublimation growth rate of ice crystals having a form other than spherical still remains practically unstudied. Approximate estimates and observations show that for supersaturations of the order of 10–12% (assuming a concentration of about 100 per cubic meter) the ice crystals may grow within 4–5 min to a size with fall velocities of up to 30–40 cm/sec; within 1–2 min the mass of a crystal will equal that of a water droplet with a radius of about 100–200 μ. This rapid growth by sublimation is the reason why even thin (about 1 km thick) ice-crystal clouds with small velocities of rising air currents can produce precipitation bands which sometimes reach the earth's surface in the form of fine, light snow or rain (examples are altostratus and cirrostratus).

Substantially different conditions prevail when the ice particle occurs in the vicinity of water droplets, as is the case in mixed clouds. Conditions here are very favorable for sublimation growth, especially if there are many more supercooled droplets than crystals in the cloud. In middle latitudes such clouds are represented by Ns and Cb. When relative humidity in the cloud decreases as a result of the sublimation of water vapor on ice particles conditions of phase equilibrium over the droplets are disrupted and the latter start evaporating, thereby adding to the supply of moisture

for crystal growth. Thus a special process of "transfer" (distillation) of water from supercooled droplets to crystals begins to operate.

TABLE 67

Critical diameters of spherical ice particles

Temperature (°)	0	-1	-5	-10	-15	-20	-30	-40
Critical diameter (10^6 cm)	∞	7.80	3.10	1.68	1.18	0.93	0.67	0.56

Table 67 gives the critical diameters for which spherical ice particles begin to grow by distillation of water from large water drops as a function of temperature.

According to the table, in a supercooled droplet cloud rapid growth by distillation of vapor from droplets will set in as soon as ice particles with a radius of the order of 10^{-6} cm appear. As a result the crystals will be able to grow to large sizes until all the droplets evaporate, the initial excess of water vapor condenses upon the crystals and the cloud has been completely transformed into an ice-crystal cloud. Calculations show that the role of this process in the initial growth of ice particles occurring in a medium containing supercooled droplets is very great.

The initial stage of ice-crystal growth by sublimation takes place far more rapidly (10–20 times) than the condensational growth of water droplets, and the point at which the process of accretion becomes dominant in the further growth of the ice crystals ($r = 50–60\mu$) is reached within a few minutes.

When considering the accretion of ice crystals it should be borne in mind that falling crystals (especially stars and plates) have a greater capture surface than droplets for the same mass; at the same time their fall velocity is lower than that of spherical particles. This accounts for the faster growth of nonspherical ice crystals by accretion and for the diversity of their shapes. However, the problem of the accretion of such particles is very difficult to solve. It may roughly be estimated that ice particles (flat prisms) grow 5–6 times more rapidly than droplets with the same mass.

Observations show that when ice particles collide with supercooled droplets the latter freeze on the crystals, a phenomenon which is expressed either in riming of the crystal or in the formation of a water film which envelops the crystal and subsequently freezes.

If the cloud contains a large number of fine droplets they will freeze almost instantaneously on the crystals without changing shape and produce a friable granular layer of ice which envelops the crystal and converts it into a pellet (strong riming). When larger drops collide with a crystal they partly spread out over its surface and form a film which on subsequent freezing produces a denser layer of ice. When large supercooled droplets freeze on an ice surface they spread out completely over the surface of the ice particle and form a continuous layer of ice (usually transparent), producing a hailstone.

For a small initial grain size and a disorderly type of fall, hailstones are nearly spherical in shape and often have a many-layered structure with alternating layers of clear and matt ice (depending on conditions under which these layers were formed).

Particles of complicated shape and considerable size may form when pellets, snow crystals or hailstones freeze together on collision. Thus particles of graupel falling in the form of small flakes of snow have diameters between 0.3 and 5 mm. The diameter of hailstones is generally several millimeters (up to 1 cm), though in certain cases it may reach 10 cm and possibly even more.

Graupel and hail fall in short-lived showers from cumulonimbus. Their duration is usually limited (less than an hour). For instance, according to observations in the Rostov region, hail showers lasting up to 5 min account for 50%, 5–20 min for 30% and 20 min and over for 20% of the total.

The area over which hail falls ranges from fractions of a kilometer to 10 km in width and is several tens of kilometers in length. This zone is frequently called the "hail path". The hail which falls may produce a layer up to 20 cm thick.

The number of days with hail depends on local conditions. In the USSR hail is frequent in the Caucasus (Georgia, Armenia) and in the southern part of Kazakhstan, where there are 8–10 days with hail per year. Less than two days with hail are recorded per annum in the Moscow district. Hail falls when air temperatures near the ground are above 0°, usually in late spring and early autumn. Hail showers are very often (but not always) accompanied by thunderstorms.

No complete theory of graupel and hail covering all complexities and conditions of formation is available. The only work done so far is the few calculations mentioned above concerning the growth of spherical ice particles in clouds of mixed structure, which have led to a revision of the role of ice particles in the formation of precipitation.

§ 8. Amounts and global distribution of precipitation

After suitable processing measurements of precipitation at meteorological stations furnish data of great practical value concerning the overall amount of water falling in different regions of the earth over more or less extensive periods of time (year, season, month). Globally the maximum yearly precipitation (on the average) is received in the equatorial zone and the minimum in high latitudes (Arctic and Antarctic) and the subtropical desert zone.

The most humid regions receive over 200 cm of precipitation annually; in certain regions (Pacific islands and elevated coasts of continents) the annual precipitation exceeds 300–400 cm. Record amounts have been reported on Waialeale peak on Kauai Island (Hawaiian Islands) and at Cherrapunji (India). There the amount of rainwater falling in a year averages 12 m and reaches 15 m or even more in certain years.

Apart from such regions of abundant precipitation, some places receive no precipitation for several years, e.g., the very dry desert areas of Chile, Peru and the Sahara. Throughout the subtropical zone of high pressure and minimum cloudiness, the total precipitation is less than 25 cm per annum.

In middle latitudes the total precipitation varies over a fairly broad range from place to place. Coastal areas generally have 75–100 cm of precipitation, inland regions 25–50 cm.

Precipitation increases with elevation above sea level (in mountains) up to a certain level, then decreases.

In the USSR maximum precipitation occurs on the southern slopes of the Glavnyi Kavkaz (Major Caucasus) range and in the so-called humid subtropics — the Black Sea coast of the Caucasus (over 200 cm yearly). Near the western border of the European USSR the annual precipitation amounts to 65–70 cm; it drops to 50–60 cm in the central districts and is somewhat lower farther east. The lowest precipitation is observed in Central Asia and in the south-eastern districts of the European USSR (to 10 cm yearly).

If the temperature of the earth's surface and adjoining layers of air is below 0°, precipitation, which will then fall chiefly in the form of snow, will produce a snow cover.

The thickness of this snow cover depends on the amount of fallen snow, its density, local relief and vegetation. The thickest snow mantles in the USSR occur on the western slopes of the Urals, in the middle and western parts of Siberia and in the Far East (lower reaches of the Amur River). In these areas the snow cover is about 1 m thick. North and south of these areas the thickness decreases. In high mountains (Caucasus, Pamir, etc.) the snow cover reaches several meters.

The density of the snow mantle varies considerably; in the USSR the mean density is about 0.2 g/cm^3 in winter and up to 0.3 g/cm^3 in spring. If the mean density of the snow ρ_{sn} and its thickness h are known one can determine the amount of water contained in the snow cover; obviously, $W = \rho_{sn} h$ g/cm^2. Part of the water reservoirs accumulated during the winter are used up in wetting the soil when the snow melts; this is very important for growing agricultural crops.

An important characteristic of precipitation is its intensity, which decreases polewards from the equatorial latitudes. The heaviest precipitation usually falls in the form of showers. Cases where the intensity reached 10 mm per minute are known to have occurred. In the USSR showers with an intensity of 5–6 mm per minute have been observed repeatedly. In the tropics a daily precipitation of up to 1000 mm is frequent (1168 mm of rain fell in one day in the Philippines in 1911). The largest daily precipitation recorded in the USSR is about 300 mm.

§ 9. Hydrologic cycle. The moisture balance at the earth's surface

Though considered separately the processes of evaporation and precipitation are very intimately related to each other; in fact, they only represent separate links in the general hydrologic cycle of the earth.

In the present era the amount of water occurring on earth in all three aggregate states may be regarded as invariant. This allows us to assume that the moisture regime of the globe as a whole is stable: the sum of precipitation falling on the surface is equal to the total amount of water evaporated.

An equation of the moisture balance similar to that of the thermal balance can be set up for any given time interval.

If the moisture balance of sea and land and also of different land regions are to be considered individually, other components of the balance in addition to evaporation and precipitation must also be considered; these include in the first place surface runoff, i.e., the amount of water removed

from the continents to the sea by rivers. Certain components, for instance the loss of water by percolation (seepage of moisture into the soil) or the exchange between land and sea through ground waters, may be disregarded in the first approximation. On the yearly average the equation of the moisture balance for the ocean surface may then be written as follows:

$$r_o = E_o - f, \tag{34}$$

where r_o is the precipitation, E_o the evaporation and f the surface runoff.

For the total land surface we write, similarly,

$$r_l = E_l + f_l \tag{35}$$

For the entire surface of the globe, disregarding the drainage into inland basins (i.e., taking $f = f_l$), we add (34) and (35) and obtain

$$r_o + r_l = E_o + E_l \tag{36}$$

Quantitative evaluation of the components of the moisture balance is greatly complicated by the lack of sufficient precipitation data for the oceans. According to present data (accurate to 10%), the total precipitation falling on the surface of the global oceans r_o yields an average of 102 cm/annum, which corresponds to $370 \cdot 10^3$ km^3 of water. The evaporation from the ocean surface, on the other hand, amounts to 113 cm/annum or, in total, $407 \cdot 10^3$ km^3. The water discharged into the oceans by rivers is 10.3 cm/annum or $37 \cdot 10^3$ km^3, which amounts to 10% of the annual precipitation over oceans

For the land surface, allowing for runoff into the sea, we have the following values: precipitation 70.0 cm/annum ($102.0 \cdot 10^3$ km^3), evaporation 44.6 cm/annum ($64.9 \cdot 10^3$ km^3) and runoff 25.4 cm/annum ($37 \cdot 10^3$ km^3). Land areas deprived of drainage into the sea have in all about $8 \cdot 10^3$ km^3 of precipitation and the same amount of evaporation.

Globally, the total annual layer of precipitation, which coincides with the total annual amount of evaporation, is 92.8 cm/annum ($370.1 \cdot 10^3 + 102.0 \cdot 10^3 = 472 \cdot 10^3$ km^3); of this in the case of precipitation 22% falls over land and 78% over sea, and in the case of evaporation 14% is from land and 86% from the sea.

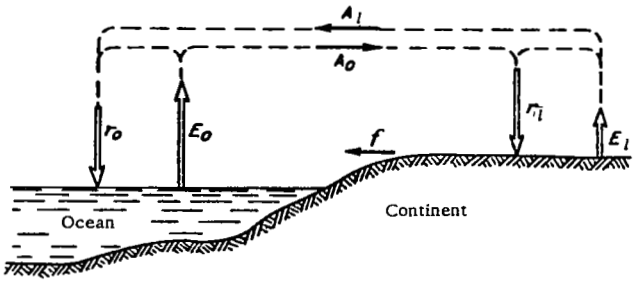


FIGURE 128. General scheme of the hydrologic cycle

A very important conclusion follows from these data: over sea less water is precipitated than evaporated, while over land, on the contrary, more moisture is precipitated than evaporated. Obviously, part of the moisture evaporated from the oceans is carried by air currents to the

continents, where it settles in the form of precipitation. But aside from this sea-to-land movement of moisture, a partial counterflow of moisture from land to sea, doubtless takes place. The overland excess of precipitation over evaporation leading to the surface runoff is a result of the mutual exchange of atmospheric moisture between oceans and continents.

Thus the general scheme of the external hydrologic cycle of the earth may be pictured as follows (Figure 128). The water evaporated from the ocean surface (E_o) is partially removed by the current A_o, just as the water evaporated from the land surface (E_l) is partly transported to the ocean by the current A_l. The difference between the currents $A_o - A_l$ determines the surface runoff f. The precipitation falling both on the ocean (r_o) and on land (r_l) consists of moisture evaporated from both ocean (E_o) and land (E_l).

In addition to the moisture exchange between land and oceans, which is called the external hydrologic cycle, the continents are the site of an internal hydrologic cycle: part of the moisture reaching the ground in the form of precipitation is evaporated there and may again become involved in the formation of precipitation. The ratio of local moisture to oceanic moisture in overland precipitation is an important question for land-improvement measures (swamp drainage, construction of irrigation systems, reservoirs, etc.).

We briefly present the fundamental relations obtained by M. I. Budyko and O. A. Drozdov for the internal hydrologic cycle.

Let us assume that a quantity of water vapor A is delivered by air currents to a land area across a unit length of its boundary. It can be set equal to $\overline{Wu}$, where $\overline{u}$ is the mean velocity of the air currents and $\overline{W}$ is the mean moisture content of the advancing air. As the air flow advances over the area in question part of this moisture will be released from it in the form of precipitation r_A, but at the same time the flow will be enriched by moisture evaporated from the land area E, part of which, in turn, may settle upon the area in the form of precipitation of local origin r_L.

The total precipitation received by the given area can be regarded as the sum of the precipitation r_A due to the intruding moisture and the precipitation r_L due to the moisture EL evaporated from this area (L being the linear dimension of the area in the direction of transport). Thus the external current, which is equal to $\overline{Wu}$ on arrival, amounts to $\overline{Wu} - r_A L$ on departure from the area; but to this must be added the stream of local water vapor, which is zero on the windward side of the boundary and $(E - r_L) L$ on the leeward side. In total this gives $\overline{Wu} - (r - E) L$

Assuming that the variation of these flows in the direction of transport takes place linearly, one can write the average external flow over the area in question as

$$\frac{\overline{Wu} + (\overline{Wu} - r_A L)}{2} = \overline{Wu} - \frac{r_A L}{2}, \tag{37}$$

while the flow of local vapor

$$\frac{0 + (E - r_L) L}{2} = \frac{(E - r_L) L}{2}. \tag{38}$$

In sum they give the total mean flow

$$\overline{Wu} - \frac{(r_A + r_L - E) L}{2} = \overline{Wu} - \frac{(r - E) L}{2}. \tag{39}$$

Since the sums of precipitation r_A and r_L are in the same ratio as the quantities of water vapor originating outside and inside the area under consideration, we can write

$$\frac{r_A}{r_L} = \frac{W\bar{u} - \frac{1}{2} r_A L}{\frac{1}{2}(E - r_L) L}, \tag{40}$$

whence approximately

$$\frac{r_A}{r_L} = \frac{1}{1 + \frac{EL}{2W\bar{u}}} \quad \text{and} \quad \frac{r_L}{r} = \frac{1}{1 + \frac{2W\bar{u}}{EL}}. \tag{41}$$

The above relations make it possible to determine what proportion of the precipitation over a given land area is due to moisture imported from outside the area and what proportion is due to moisture supplied by evaporation from the area. This relationship is expressed by means of the coefficient of water circulation k, which is equal to the ratio of the total sum of precipitation r to the amount of precipitation of external (advective) origin r_A

$$k = \frac{r}{r_A} = 1 + \frac{EL}{2W\bar{u}}. \tag{42}$$

The coefficient k obviously depends on the dimensions of the area under consideration, increasing with the latter; at the same time it depends on the evaporation E and on conditions of circulation (velocity $\bar{u}$) over the given area. It is therefore different under different physico-geographic conditions and varies with time.

Calculations carried out by many authors show that precipitation due to local water vapor amounts to only a very small part of the total precipitation on the area in question. Thus for a land area as extensive as the European USSR, according to Budyko and Drozdov, the yearly mean of k amounts to only 1.13, i.e., only 13% of the total precipitation over this area is due to water vapor contributed by local evaporation. The values of k are different in different months: from 1.05 in October-November to 1.24 in April. Calculations by K. I. Kashin and Kh. P. Pogosyan for the Ob basin also show that local evaporation from the land surface amounts to only about 8% of the total moisture contained in the air and passing over this region.

§ 10. Artificial stimulation of cloud and fog

Attempts to produce rainfall from clouds, especially during drafts, date back very far and involve a wide variety of methods; in all cases, however, the attempts either ended in failure or gave doubtful results.

The first report of a successful experiment was published in 1931 by Veraart (Holland). In his report it was stated that rain fell over an area of about 8 km^2 after 1.5 tons of dry ice had been scattered in cloud from an aircraft. The scientific explanation of the result was erroneous. (Veraart assumed that electrical charges on particles of CO_2 were the determining

factor.) Repetition of these experiments in other countries did not produce the desired effect and thus for a certain time (over ten years) all such testing was abandoned and Veraart's result was judged to be accidental.

The opinion of many eminent scientists of that period was that the artificial production of rainfall would require the expenditure of an enormous amount of energy commensurate with the work done in nature in the vertical ascent of large masses of air. Hence, the reasoning went, solution of the problem of artificial rainfall was utopian.

However, according to a different view which was also put forward, the solution to the problem should be sought not in operation on large-scale phenomena but rather in stimulation of the elementary microphysical processes taking place in clouds. Arguments were advanced to show that in certain states of the cloud a small additional impulse would be enough to stimulate or inhibit the natural course of the process of formation of precipitation. From this it was concluded that for successful stimulation of clouds it was necessary to find the appropriate methods and, most important of all, favorable conditions.

In the Soviet Union work on active cloud stimulation was begun with the support of Party and Government in 1930–1931 and was developed under the direction of Prof. V. N. Obolenskii at the Institute of Experimental Meteorology (during the war this institute was incorporated into the Main Geophysical Observatory).

Decisive and conclusive advances were made almost simultaneously and independently in the USA (by Schaefer in 1946) and in the USSR (Piotrovich et al. in 1946-1947). It was discovered that when small CO_2 particles were introduced into a chamber containing supercooled droplets they created a very large number of ice particles (up to 10^{14}–10^{16} crystals per gram of CO_2 depending on the temperature of the environment).

Positive results were obtained when particles of CO_2 were dropped into supercooled clouds from an aircraft. Within 10–15 min the ice crystals formed in these clouds had grown large enough to fall. After stimulation clearings soon appeared in the clouds and (occasionally) snow or rain fell from the clouds. These experiments made Veraart's results comprehensible: he had introduced the dry ice into supercooled clouds, a circumstance which was not considered in later experiments.

The next step in the solution of the problem of stimulation of supercooled cloud were the results obtained by Vonnegut (USA). In his experiments substances producing smoke particles with a structure isomorphic to ice (e.g., silver iodide) were vaporized in a supercooled fog (cloud) and led to the appearance of a large number of ice crystals. Silver iodide smoke proved just as effective as CO_2 when introduced into supercooled clouds at temperatures below -5, $-7°$.

Thus the combined efforts of many scientists in different countries led to the discovery of a method for stimulating supercooled cloud. This discovery has great scientific and practical values and is rightly classed among the outstanding achievements of contemporary science and technology.

A great number of successful experiments on the dissipation of cloud and fog (very important for takeoff and landing of aircraft) and the production of precipitation from clouds have been carried out in various countries in the last ten to fifteen years.

These experiments may be divided into two groups according to the physical principles on which they are based:

1) experiments in supercooled clouds. In this case the methods of stimulation are meant to induce, by artificial means, the formation of ice crystals in a medium consisting of supercooled droplets of water;

2) experiments on the stimulation of cloud and fog at temperatures above 0°. In these experiments particles introduced into a cloud are made to grow until they reach a size at which they begin to fall.

In the stimulation of supercooled cloud and fog, conditions are created which essentially reproduce the natural process of formation of precipitation from mixed clouds (according to the Bergeron-Findeisen mechanism); this process has been discussed earlier. Owing to the difference between the buoyancy of vapor over water and over ice, crystals grow while drops evaporate.

Dry ice is widely used to induce the formation of the solid phase in supercooled droplet clouds.

Due to loss of the heat of evaporation the temperature of the dry ice particles drops to −78°. The air surrounding a dry ice particle is chilled, leading to a supersaturation strong enough for minute particles consisting of a few tens of water molecules to appear and develop. At temperatures below −40°, according to V. Ya. Nikandrov, the molecule aggregates yield ice-crystal embryos. The total number of embryos formed upon evaporation of one gram of CO_2 can be found (according to A. D. Solov'ev) from the following considerations.

The evaporation of particles of dry ice requires the expenditure of the heat $Q_{CO_2}=150$ cal/g. The dry ice particle obtains this heat by chilling a certain volume of air and by using the heat of sublimation released in the formation of ice crystals. Hence the total heat balance amounts to

$$Q_{CO_2} = V\rho c_p \Delta T + V\rho L \Delta q, \tag{43}$$

where V is the volume of the chilled air, ρ its density, ΔT the temperature difference caused by chilling, L the heat of sublimation of water vapor and Δq the amount of moisture condensed per unit mass of cooled air.

Obviously, the quantity Δq should be equal to the total mass of the embryos formed in this process, i.e.,

$$N_{em} m_{em} = \Delta q V \rho, \tag{44}$$

where N_{em} and m_{em} are the number and mass of the embryos.

Then from (43) and (44) we obtain

$$Q_{CO_2} = V\rho \left(c_p \Delta T + L \Delta q \right) = \frac{N_{em} m_{em}}{\Delta q} \left(c_p \Delta T + L \Delta q \right) \tag{45}$$

or

$$N_{em} = \frac{Q_{CO_2} \Delta q}{m_{em} (c_p \Delta T + L \Delta q)}. \tag{46}$$

We will assume for simplicity that the ice embryos are spherical. Then the radius of the "viable" embryos which may exist and develop in the cloud should satisfy the condition which follows from Kelvin's formula

$$r_{em} \geqslant \frac{2\sigma}{\rho_{em} R_{vap} T \ln \frac{E_r}{E_s}}, \tag{47}$$

where σ is the surface energy (ice-air); approximately $\sigma \approx 82$ erg/cm^2.

In the given case the ratio $\frac{E_r}{E_s}$ which allows for the influence of the curvature of the surface can be replaced by the ratio $\frac{E_w}{E_i}$, since the crystals

are in an environment which is supersaturated over ice. The ratio $\frac{E_w}{E_i}$ depends on temperature (cf. Chapter 16, § 2). Thus for $T = -10°$, $\frac{E_w}{E_i} \approx 1.1$. Then $m \approx 10^{-17}$ g and $N_{em} \approx 10^{16}$ crystals per gram of carbon dioxide.

This gives the maximum number of crystals which may be formed for complete exploitation of the chilling power of dry ice. Since (according to Nikandrov) viable ice embryos are formed by complexes of vapor molecules at temperatures below $-40°$, the above values of N_{em} must be reduced by two to three orders (according to laboratory data N_{em} amounts to $10^{11}-10^{12}$ crystals per gram of CO_2)

When particles of dry ice are introduced into a supercooled cloud they give rise to a large number of ice crystals which travel inside the cloud over considerable distances (several kilometers) owing to turbulent mixing. If the concentration of crystals is very high the droplet cloud will become an ice-crystal cloud.

In order for crystals to grow to the critical size for which release is possible (that is, in order for them to reach the peak of the trajectory in the cloud), a certain optimum ratio between the number of crystals and droplets which depends on the velocities of the vertical air motions must be achieved; it then becomes possible to obtain rain or snow from the cloud In practice these velocities are not always known in experiments and usually 1–2 kg of CO_2 are used up for every kilometer of flight path.

Thus the problem of stimulating supercooled clouds has been solved in principle; what remains to be perfected are the technical and practical aspects.

We now turn to the stimulation of cloud and fog at positive temperatures. Here, as before, a distinction must be drawn between two independent problems: firstly, the dissipation of cloud and fog, or rather the production of clearing inside them, and, secondly, the stimulation of precipitation, chiefly from convective clouds.

We first consider the former. Several procedures for dissipating fog and low stratus cloud were suggested at various times. Of these two have very sound physical bases.

1. The thermal method, intended chiefly for fog. The principle on which it rests is very simple. When air is heated to a temperature above the dew point the droplets of fog evaporate and the fog dissipates. Various kinds of heaters (oil burners, coal, etc.) were used for this purpose during the second world war and earlier. This method requires a large expenditure of heat: for a liquid-water concentration in the fog $a_{liq} = 0.5\ g/m^3$ about $300\ cal/m^3$ are necessary to evaporate the droplets. To warm one cubic meter of air with a mass of roughly 1.2 kg (heat capacity $c_p = 0.24$ kcal/deg) by one degree one must expend $q = mc_p\Delta T \approx 300\ cal/m^3$. Since to scatter a fog one usually needs to warm the air by several degrees, the expenditure in heat without allowing for mixing amounts to about $10^3\ cal/m^3$. To maintain a clearing 100×100^2 in cross section for a wind speed $v = 5$ m/sec one must burn not less than 100 tons of oil per hour. Therefore the thermal method is used only in exceptional cases.

2. Method based on the utilization of hygroscopic substances. This was proposed in 1935–1937 by V. A. Fedoseev and subsequently patented in the USA in 1939 by Radford and Houghton. The principle of this method can be explained as follows. As is known, the buoyancy of vapor over a solution of hygroscopic substance decreases with concentration according to

Raoult's law (cf. Chapter 18, § 2). When hygroscopic particles are introduced into a fog (cloud), at the initial stage the particles grow by condensation of water vapor. Absorption of moisture may lead to lowering of the relative humidity to values below 100%; the fog (cloud) droplets will then begin to evaporate. If the hygroscopic particles reach fall sizes in this process, the fog (cloud) should dissipate.

The optimum expenditure of hygroscopic substance (without considering ascending air motions) is easily determined. We assume that for the fog droplets to evaporate it is sufficient to reduce the relative humidity by 10%; then the appropriate concentration of solution for, say, NaCl is 125 g of NaCl per liter of water, i.e., one gram of NaCl when rationally utilized may condense 8 grams of water. If the liquid-water content of the fog $a_{liq} = 0.5-0.8\ g/m^3$, $T = 0°$ and $a_{vap} = 4.8\ g/cm^3$ then the dissipation of one cubic meter of fog will require the following amount of hygroscopic substance

$$M = \frac{0.1\, a_{vap} + a_{liq}}{8} \approx 0.1 \text{ g NaCl/m}^3.$$

Technical difficulties and the large amounts of hygroscopic substance involved hamper practical application of this method.

At present acoustic, high-frequency, electrical, chemical (introduction of surface-active substances) and other means of dissipating fog and low cloud are being studied under laboratory and field conditions.

Let us now turn to the second problem — that of stimulating precipitation from "warm" liquid clouds. Hygroscopic substances are also used for this purpose. The physical foundations of this method are as follows. In middle latitudes even thick cumulus clouds do not usually yield precipitation, since the cloud droplets cannot grow to fall sizes. If hygroscopic particles larger than natural condensation nuclei are introduced artificially into the base of such clouds, these particles (or rather the droplets of solution) will soon grow by condensation to the "threshold" at which the accretion mechanism begins to operate. On further growth the resulting drops will be able to reach the peaks of their trajectory and subsequently fall as rain.

Initial sizes required of the hygroscopic particles depend on the velocity of the ascending motions and the vertical thickness of the cloud. Calculations performed with the formulas in sections 5 and 2 of Chapter 18 show that this method of stimulation is most effective in the case of dense cumulus clouds.

The efficiency of the method is easily estimated from the ratio of the initial radius r_{in} of the seeded particles to their final radius r_{fin}, since the ratio of the mass of substance introduced to the mass of precipitation released corresponds roughly to the ratio of the cubes of their radii $\left(\frac{r_{fin}^3}{r_{in}^3} = \eta\right)$. Thus for a vertical velocity inside the cloud $w \approx 1$ m/sec, droplets of saturated NaCl solution with $r_{in} = 20\mu$ will reach the peak of the trajectory at the height $z_{peak} \approx 1$ km and will ultimately attain $r_{fin} \approx 200\mu$; then $\eta = 10^3$. For $w = 5$ m/sec and $z_{peak} \approx 4$ km particles with an initial size of 20μ will grow to 2000μ and $\eta = 10^5 - 10^6$.

This means that in the former case ($w = 1$ m/sec) about 1 ton of water can be obtained for one kilogram of substance (NaCl) while in the latter cas ($w = 5$ m/sec) up to 10^3 tons of water can be obtained for the same.

In the tropics where clouds with a high liquid-water content are vigorously

developed along the vertical seeding with large water drops to create a "chain reaction" (after Langmuir) can find applications.

It is important to note that this method becomes inefficient in the case of weakly developed, thin stratiform clouds. Thus for $w = 0.1$ m/sec and $r_{in} = 20\mu$, the substance should be introduced into the upper part of the cloud; then for a cloud thickness of 0.5 km the droplets will grow to $r_{fin} \approx 50\mu$ and η will be only 10–15.

A difficult problem which always arises when evaluating the results of tests is that of proving that the effect obtained was produced artificially and not by the natural development of the process. This is done by various means including: radar observations to ascertain the locus of operation and the changes taking place in the state of the cloud; ground observation of the quantity of precipitation falling outside and inside the zone of operation; and, recently, statistical climatological methods in which precipitations inside and outside the zone of operation are compared on the average over several years over a comparatively large area.

The results obtained open up wide prospects for cloud stimulation for such practical purposes as: redistribution of precipitation; prevention of hail and downpours; dissipation of cloud and fog; modification of climatic conditions, and so forth.

New striking advances in this field are surely in store for us in the near future.

Part Five
AIR MOTIONS

Air motions differing widely in intensity and extension are constantly observed in the atmosphere. The immediate cause of atmospheric motion is the uneven distribution of pressure due to processes of heat transfer. The study of atmospheric motion provides the key to the fundamental regularities governing weather and climate and is the main aim of dynamic meteorology, an aim which it pursues by theoretical methods based on the equations of fluid mechanics and thermodynamics. Below we shall confine ourselves to certain of the fundamental ideas associated with this extensive branch of meteorology.

Chapter 21

GENERAL REGULARITIES OF ATMOSPHERIC MOTION

§ 1. Forces in the atmosphere

In our study of atmospheric motion we will begin with a consideration of the forces under the influence of which this motion takes place.

The forces in the atmosphere can be classed into two groups: body (volume) and surface forces. The first group comprises those forces which act upon each element of mass (or volume) irrespective of whether other air particles are present near the element in question. Examples are the force of gravity and the inertial forces; the latter include the deflective force of the earth's rotation and the centrifugal force.

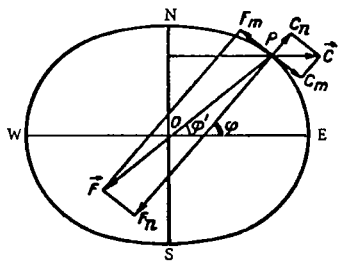


FIGURE 129

Surface forces are the forces of interaction of a certain volume of air with the surrounding medium. These forces act upon the surface particles of the volume under consideration, and include the pressure force and the viscous forces.

Let us consider these forces one by one.

The force of gravity g is the difference between the normal component of the force of attraction to the earth's center $\vec{F}$ and that of the centrifugal force $\vec{C}$ directed along the radius-vector of rotation. Denoting these components by F_n and C_n, we obtain for the force of gravity g per

unit mass (acceleration of gravity)

$$g = F_a - C_a. \tag{1}$$

From Figure 129, we easily see that

$$F_a = k\frac{M}{a^2}\cos\gamma \quad \text{and} \quad C_a = C\cos\varphi = \omega^2 a\cos\varphi'\cos\varphi, \tag{2}$$

where $k = 6.67 \cdot 10^{-8}\ \text{cm}^3\ \text{g}^{-1} \cdot \text{sec}^{-2}$ is the gravitational constant, M is the mass of the earth, a is the distance of the point P under consideration from the earth's center, $\omega = 7.29 \cdot 10^{-5}\ \text{sec}^{-1}$ is the angular velocity of rotation of the earth, φ is the geographic latitude, φ' the geocentric latitude, and $\gamma = \varphi - \varphi'$. In view of the smallness of γ one can set $\cos\gamma = 1$ and $\varphi' = \varphi$ and replace a by the mean radius of the earth R. Then expression (1) for the acceleration of gravity takes the form

$$g = k\frac{M}{R^2} - \omega^2 R\cos^2\varphi. \tag{3}$$

The acceleration of gravity can be expressed in terms of the geopotential $W\left(g = -\frac{\partial W}{\partial n}\right)$; it varies along the level surface and with elevation above sea level. Its mean value at latitude $\varphi = 45°$ and sea level is $g = 980.62\ \text{cm/sec}^2$.

The centrifugal force $\vec{C}$ is greatest at the equator, where it is $C_0 = 3.4\ \text{cm/sec}^2$; it amounts to a very small fraction of the force of attraction $\vec{F}$; at the equator $\frac{C_0}{F} = \frac{1}{289}$.

Deflective force of the earth's rotation. It is well known from mechanics that every body moving over the earth's surface experiences a relative acceleration (Coriolis acceleration) due to the fact that the earth rotates about its axis. This force is inertial; it acts only on moving masses and shows up only when motion is considered in a coordinate system rigidly fixed to the rotating earth.

In the general case the magnitude and direction of the deflective force are given by twice the vector product of the angular velocity of rotation by the velocity of the body with respect to the earth. Denoting the vector of the deflective force by $\vec{K}$, the angular velocity by $\vec{\omega}$ and the velocity of the body by $\vec{V}$, we obtain

$$\vec{K} = -2\left[\vec{\omega}\,\vec{V}\right]. \tag{4}$$

The components of this force along the coordinate axes are obviously

$$\left.\begin{aligned} K_x &= 2v\omega_z - 2w\omega_y \\ K_y &= 2w\omega_x - 2u\omega_z \\ K_z &= 2u\omega_y - 2v\omega_x \end{aligned}\right\} \tag{5}$$

where u, v, w and ω_x, ω_y, ω_z are respectively the components of $\vec{V}$ and $\vec{\omega}$ along the coordinate axes.

In the northern hemisphere this force acts to the right, and in the southern hemisphere to the left, of the direction of movement.

If we take the so-called standard coordinate system, directing the P-axis upward along the vertical, the z-axis from west to east along the parallel of latitude, and the y-axis from south to north along the meridian, then

$\omega_x=0$, $\omega_y=\omega\cos\varphi$, $\omega_z=\omega\sin\varphi$, and the expressions (5) become

$$\left.\begin{aligned} K_x &= 2v\omega\sin\varphi - 2w\omega\cos\varphi \\ K_y &= -2u\omega\sin\varphi \\ K_z &= 2u\omega\cos\varphi \end{aligned}\right\}. \qquad (5')$$

The vertical component K_z depends only on the velocity component u acting eastwards or westwards and acts either in the direction of gravity (if $u<0$) or in the opposite direction $(u>0)$. If one compares its numerical value with the value of g one finds that K_z is four orders (10,000 times) smaller than g. This component is therefore of no particular significance for motions taking place in the atmosphere and is usually not considered. The horizontal components cannot, however, be disregarded.

Since in the atmosphere the horizontal components u and v are usually many times greater than the vertical component w, in many cases the term $2w\omega\cos\varphi$ in the first equation of (5') can be disregarded by comparison with the term $2v\omega\sin\varphi$. Thus instead of (5') we will have

$$\left.\begin{aligned} K_x &= 2v\omega\sin\varphi \\ K_y &= -2u\omega\sin\varphi \end{aligned}\right\} \qquad (6)$$

and the magnitude of the horizontal component of $\vec{K}$ will be

$$\left|\vec{K}\right| = K = \sqrt{K_x^2+K_y^2} = 2c\omega\sin\varphi, \qquad (7)$$

where $c=\sqrt{u^2+v^2}$ is the horizontal component of the velocity.

To illustrate the magnitude of the deflective force K we present its numerical values for a unit mass at different latitudes, assuming $c=$ $=10\,\text{m/sec}$ (Table 68).

TABLE 68

φ (°)	10	30	50	70	90
K (cm/sec^2)	0.025	0.073	0.112	0.137	0.146

From this it is evident that K is small in comparison with g. But since this force acts at right angles to the direction of the velocity, it appreciably modifies the direction, though without changing the magnitude of the latter.

The pressure force G is the force arising as a result of nonuniform distribution of pressure.

We recall that, from aerodynamics, the conditions of equilibrium of an air mass under the influence of gravity alone are (for an air mass equal to unity):

$$X-\frac{1}{\rho}\frac{\partial p}{\partial x}=0,\quad Y-\frac{1}{\rho}\frac{\partial p}{\partial y}=0,$$

$$Z-\frac{1}{\rho}\frac{\partial p}{\partial z}=0, \qquad (8)$$

where X, Y and Z are the components of the force of gravity and $-\frac{1}{\rho}\frac{\partial p}{\partial x}$, $-\frac{1}{\rho}\frac{\partial p}{\partial y}$, $-\frac{1}{\rho}\frac{\partial p}{\partial z}$ those of the resultant hydrostatic pressure per unit mass.

According to these equations, for air to be in quilibrium it is necessary that the force of gravity be balanced by the hydrostatic pressure experienced by the given mass from the surrounding medium.

Placing the z-axis vertically and the plane xy horizontally, we have

$$X=Y=0,\quad Z=-g,$$

where g is the acceleration of gravity. Then the system of equations (8) becomes

$$-\frac{1}{\rho}\frac{\partial p}{\partial x}=0, \quad -\frac{1}{\rho}\frac{\partial p}{\partial y}=0, \quad -g-\frac{1}{\rho}\frac{\partial p}{\partial z}=0. \tag{9}$$

From this it is seen that for air to be in equilibrium on a horizontal plane the atmospheric pressure must be constant on this plane, i.e., that

$$\frac{\partial p}{\partial x}=\frac{\partial p}{\partial y}=0.$$

As we know, a surface (here a horizontal plane) on which the pressure is constant at all points is called isobaric. But owing to the uneveness of the pressure distribution in the atmosphere, isobaric surfaces have a complex shape and are inclined with respect to the level surfaces. Their intersection with the horizontal surface corresponding to a given level gives a complex system of isobars. The pressure change per unit length along the normal n to an isobar (in the direction of decreasing pressure) is the horizontal pressure gradient $-\frac{\partial p}{\partial n}$. In practice the length unit is taken to be the length of 1° of longitude (111.2 km), and the pressure is expressed in millibars; thus the horizontal gradient is determined in mb/1°, and $1\ mb/1^\circ = 0.9 \cdot 10^{-4}\ dyn/cm^3$.

Thus the vector of the pressure force is given by

$$\vec{G}=-\frac{1}{\rho}\operatorname{grad} p, \tag{10}$$

and its components along the coordinate axes are

$$G_x=-\frac{1}{\rho}\frac{\partial p}{\partial x}, \quad G_y=-\frac{1}{\rho}\frac{\partial p}{\partial y}, \quad G_z=-\frac{1}{\rho}\frac{\partial p}{\partial z}. \tag{11}$$

The horizontal component of G

$$G=-\frac{1}{\rho}\frac{\partial p}{\partial n}, \tag{12}$$

is the fundamental force responsible for horizontal motions of air (wind).

Viscous forces (force of friction) arise when different volumes of air have different velocities.

If we regard air motion as the motion of a viscous liquid along a plane wall, experiment shows that as one moves away from the wall the velocity of the liquid increases. Then the tangential viscous stress τ arising between continuous layers of the liquid moving at different speeds turns out to be proportional to the velocity change in the direction perpendicular to the wall. Therefore $\vec{\tau}$ can be defined by

$$\vec{\tau}=\mu\frac{\partial \vec{v}}{\partial z}, \tag{13}$$

where μ is the dynamic (molecular) viscosity; the z-axis is directed at right angles to the wall. (Formula (13) can be obtained easily for an ideal gas from purely theoretical considerations.) The numerical value of the dynamic viscosity of air amounts to $1.7 \cdot 10^{-4}$ g/cm·sec at $t = 0°$ and $2.2 \cdot 10^{-4}$ g/cm·sec at $t = 100°$. However, when studying actual air motions, owing to their turbulent character the quantity μ should be regarded as the coefficient of

turbulent exchange. As we noted earlier, the coefficient of turbulent exchange is tens and hundreds of thousands of times greater than the dynamic viscosity.

Let us denote the coefficient of turbulent exchange by A. Then instead of (13) we have

$$\vec{\tau}_1 = A\frac{\partial \vec{v}}{\partial z} = k\rho\frac{\partial \vec{v}}{\partial z}, \tag{14}$$

where k is the coefficient of turbulence.

We note that the velocity $\vec{v}$ in the above has the significance of an average velocity.

Making use of formula (14), let us now obtain an expression for the viscous force — internal friction — $\vec{R}$. Consider the volume element of liquid represented in Figure 130. Let the areas of the upper and lower surfaces of the volume element be unity and its height be dz. Then the viscous stresses acting on the upper and lower surfaces are respectively $\vec{\tau}+\frac{\partial \vec{\tau}}{\partial z}dz$ and $\vec{\tau}$. Their resultant is $\frac{\partial \vec{\tau}}{\partial z}dz$. This is the viscous force acting on the volume element of liquid in question.

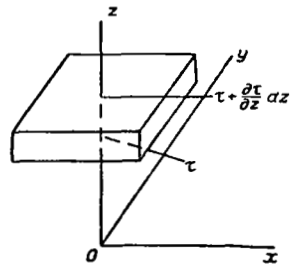


FIGURE 130

The viscous force per unit mass will be

$$\vec{R} = \frac{1}{\rho}\frac{\partial \vec{\tau}}{\partial z}, \tag{15}$$

or, in view of (14),

$$\vec{R} = \frac{1}{\rho}\frac{\partial}{\partial z}\left(A\frac{\partial \vec{v}}{\partial z}\right) = \frac{\partial}{\partial z}\left(k\frac{\partial \vec{v}}{\partial z}\right). \tag{16}$$

In the general case, where the velocity does not change in the direction of the z-axis alone, formula (16) is replaced by the following

$$\vec{R} = \frac{\partial}{\partial z}\left(k_z\frac{\partial \vec{v}}{\partial z}\right) + k_0\left(\frac{\partial^2 \vec{v}}{\partial x^2} + \frac{\partial^2 \vec{v}}{\partial y^2}\right), \tag{16'}$$

where k_z and k_0 are the coefficients of turbulence in the vertical and horizontal directions respectively; the latter is assumed to be independent of x and y.

If one considers the combined effect of molecular and turbulent friction, the tangential stress can be expressed as the sum

$$\vec{\tau}_1 = (A + \mu)\frac{\partial \vec{v}}{\partial z}, \tag{17}$$

whence the force of friction will be

$$\vec{R}_1 = \frac{1}{\rho}\frac{\partial \vec{\tau}_1}{\partial z}. \tag{18}$$

The centrifugal force arises when motion takes place along a curvilinear trajectory. If the radius of curvature of this trajectory is r and the velocity at a certain point is v then the centrifugal force at this point, as is well known, will be

$$\vec{C} = \frac{v^2}{r}. \tag{19}$$

§ 2. Equations of motion

As is known from mechanics, the equations of motion of a moving mass of air (per unit mass) can be written as follows:

$$\frac{du}{dt} = X, \quad \frac{dv}{dt} = Y, \quad \frac{dw}{dt} = Z, \tag{20}$$

where u, v and w are the components of the velocity vector $\vec{c}$ and X, Y, and Z those of the resultant of all active forces, including volume and surface ones.

Recalling the expressions obtained in § 1 for the forces acting upon a unit mass, instead of (20) we write the following equations of motion of an incompressible fluid on a rotating earth:

$$\left.\begin{aligned}
\frac{du}{dt} &= -\frac{1}{\rho}\frac{\partial p}{\partial x} + (2\omega_z v - 2\omega_y w) + \frac{\partial}{\partial z}\left(k_z \frac{\partial u}{\partial z}\right) + \\
&\qquad + k_0\left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right] \\
\frac{dv}{dt} &= -\frac{1}{\rho}\frac{\partial p}{\partial y} - 2\omega_z u + \frac{\partial}{\partial z}\left(k_z \frac{\partial v}{\partial z}\right) + k_0\left[\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right] \\
\frac{dw}{dt} &= -\frac{1}{\rho}\frac{\partial p}{\partial z} + 2\omega_y u + \frac{\partial}{\partial z}\left(k_z \frac{\partial w}{\partial z}\right) + k_0\left[\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right]
\end{aligned}\right\} \tag{21}$$

We note here that when studying motions in the atmosphere one is dealing not with the velocity change of a moving particle of air $\frac{dc}{dt}$ but rather with the velocity change observed at a certain specified point. Hence the necessity of expressing the individual (or total) derivative $\frac{dc}{dt}$ in terms of the derivatives of the velocity referred to a definite point in space (local derivatives).

In this case, for, say, the velocity component $u(x, y, z, t)$, which is a function not only of time but also of the position of the moving particle, the individual derivative with respect to time is

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x}\frac{dx}{dt} + \frac{\partial u}{\partial y}\frac{dy}{dt} + \frac{\partial u}{\partial z}\frac{dz}{dt},$$

but since

$$\frac{dx}{dt} = u, \quad \frac{dy}{dt} = v, \quad \frac{dz}{dt} = w,$$

we have

$$\frac{du}{dt} = \frac{\partial u}{dt} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}$$

and similarly

$$\frac{dv}{dt} = \frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z},$$

$$\frac{\partial w}{dt} = \frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}. \qquad (22)$$

The system of equations of motion (21) must be supplemented by the so-called equation of continuity, which is an expression of the law of conservation of mass. To obtain it we determine the influx of air (mass) per unit time into the volume element $dv = dx\,dy\,dz$. The net influx along the z-axis is

$$\left[\rho u + \frac{\partial(\rho u)}{\partial x}dx\right]dy\,dz - \rho u\,dy\,dz = \frac{\partial(\rho u)}{\partial x}dx\,dy\,dz.$$

Analogous expressions can be obtained for the other coordinate axes As a result the total influx of mass throughout the volume in question is

$$\left[\frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z}\right]dx\,dy\,dz = \operatorname{div}\rho\vec{c}\,dx\,dy\,dz.$$

This expression gives the mass of air flowing out of the volume in question per unit time. Owing to the law of conservation of matter this flow should be compensated by a lowering of the density of air in the volume in question, which can be written as

$$\frac{\partial\rho}{\partial t}dx\,dy\,dz.$$

Setting the last two expressions equal, we obtain the following equation of continuity:

$$\frac{\partial\rho}{\partial t} + \operatorname{div}\rho\vec{c} = \frac{\partial\rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0, \qquad (23)$$

here, as one can show by comparing the order of magnitude of the individual terms, the first term can be disregarded by comparison with the remaining terms, i. e., one may write

$$\operatorname{div}\rho\vec{c} \approx 0. \qquad (24)$$

The system of equations (21) and (23) should be supplemented by the equation of heat influx (Chapter 12, § 5) and by the equation of state of moist air.

Thus, the system of equations of motion is highly complex. Its general solution is therefore fraught with difficulties and has not been obtained so far.

The fundamental equations of motion (21) can be simplified in various ways, for instance by estimating the order of magnitude of the individual terms appearing in these equations. Retaining terms of the order of 10^{-4}–10^{-3} in the first two equations of (21) and only terms of the order of 10^{1} in the third equation, one can rewrite the system as follows

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + \qquad (25)$$
$$+ 2\omega_z v + \frac{\partial}{\partial z}\left(k_z\frac{\partial u}{\partial z}\right)$$

$$\frac{dv}{dt} = \frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial y} - \\ -2\omega_z u + \frac{\partial}{\partial z}\left(k_z \frac{\partial v}{\partial z}\right) \qquad (25)$$

$$\frac{dw}{dt} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g = 0$$

the third of the above equations can clearly be reduced in many cases to the fundamental equations of hydrostatics.

§ 3. Trajectories and streamlines

The system of equations of motion (25) given in the previous section can be used to investigate atmospheric motion in either of two ways. One may, for instance, confine oneself to solving this sytem of equations for the unknown velocity components u, v, and w. In this case one will obtain the velocity field. On the other hand, one may set oneself the far more complicated task of determining the motion of individual air particles. In both cases it is useful to represent the results of integration of the equations of motion in graphical form. This kind of graphical representation is also very valuable in analyzing observational data.

When studying the field of air flows a very clear way of characterizing it is by means of so-called streamlines. Streamlines are lines the tangents to which coincide at every point with the direction of motion at the given instant. From this definition, the differential equations of streamlines can be written as

$$\frac{dx}{u(x, y, z, t)} = \frac{dy}{v(x, y, z, t)} = \frac{dz}{w(x, y, z, t)}. \qquad (26)$$

Thus streamlines characterize the velocity field at any definite instant of time.

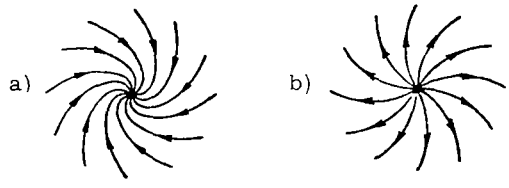


FIGURE 131. Points of convergence (a) and divergence (b) of streamlines

Aside from streamlines, one can construct and study the trajectories of particles, which give the variation with time of the position of individual particles and are described by equations of the same form as (24); in this case, however, the time is not a parameter (as for the streamlines) but an independent variable.

Naturally, if the motion is steady, i.e., if the velocity components do not depend on time, the equations of the streamlines and those of the trajectories will become identical. In this case the trajectories and streamlines will coincide. In unsteady motion, however, the trajectories and streamlines may be substantially different.

Isotachs — lines of equal wind speed — and also isogons — lines of same direction of the wind-speed vector — are frequently used for graphical representation of the velocity field. After constructing the isogons one can then draw streamlines.

In meteorology atmospheric motions characterized by the presence of convergence or divergence of the streamlines present particular interest.

Convergence of streamlines at a certain point (Figure 131a) is observed in low-pressure regions at the earth's surface. Conversely, high-pressure regions display divergence of the streamlines from a certain center (Figure 131b).

In the presence of two air currents, convergence or divergence of the streamlines may occur between the currents along certain lines (lines of convergence or divergence). These lines are illustrated schematically in Figure 132.

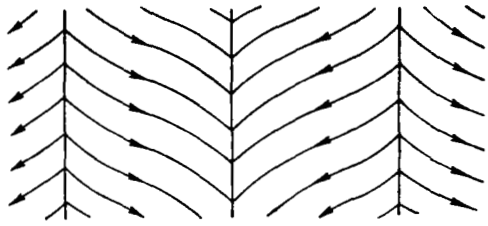


FIGURE 132. Lines of divergence and convergence

Convergence and divergence constantly occur when air currents move over dissected terrain, where the currents narrow or broaden under the influence of the obstacles they encounter; changes in wind speed are also observed occasionally

§ 4. Frictionless stationary motion

Let us consider the case of air motion in a horizontal plane in the absence of friction. This condition may be regarded as fulfilled at altitudes greater than 500–1000 m, where the influence of the force of friction is so small that it can be disregarded. Let there be a steady horizontal motion of air at these altitudes. Such a flow (in the absence of friction) is called gradient wind. Let us consider the two simplest particular cases of gradient wind: a) wind for rectilinear isobars and b) wind for circular isobars.

a) Geostrophic wind. In this case, owing to the constancy of the velocity $\left(\frac{du}{dt}=\frac{dv}{dt}=0\right)$, instead of (25) we obtain the following system of equations:

$$\left.\begin{aligned} 0 &= -\frac{1}{\rho}\frac{\partial p}{\partial x} + 2\omega_z v_g \\ 0 &= -\frac{1}{\rho}\frac{\partial p}{\partial y} - 2\omega_z u_g \\ 0 &= -\frac{1}{\rho}\frac{\partial p}{\partial z} - g \end{aligned}\right\} \qquad (27)$$

where the components of the velocity vector $\vec{c_g}$ along the coordinate axes are denoted by u_g and v_g.

The first two equations, called geostrophic relations, determine the geostrophic wind. They show that the geostrophic wind will be perpendicular to the pressure gradient. Indeed, multiplying the first equation in (27) by u_g and the second by v_g and adding the two, we have

$$u_g \frac{\partial p}{\partial x} + v_g \frac{\partial p}{\partial y} = 0, \tag{28}$$

but since

$$u_g = c_g \cos(c_g, x), \quad v_g = c_g \sin(c_g, x),$$

and

$$\frac{\partial p}{\partial x} = \frac{\partial p}{\partial n} \cos\left(\frac{\partial p}{\partial n}, x\right), \quad \frac{\partial p}{\partial y} = \frac{\partial p}{\partial n} \sin\left(\frac{\partial p}{\partial n}, x\right),$$

we have

$$u_g \frac{\partial p}{\partial x} + v \frac{\partial p}{\partial y} = c_g \frac{\partial p}{\partial n}\left[\cos(c_g, x) \cos\left(\frac{\partial p}{\partial n}, x\right) + \right.$$

$$\left. + \sin(c_g, x) \sin\left(\frac{\partial p}{\partial n}, n\right)\right] = c_g \frac{\partial p}{\partial n} \cos\left(c_g, \frac{\partial p}{\partial n}\right) = 0 \tag{28'}$$

or

$$\cos\left(c_g, \frac{\partial p}{\partial n}\right) = 0, \tag{29}$$

i.e., the directions of c_g and $\frac{\partial p}{\partial n}$ are perpendicular to each other.

Thus subject to the above restrictions — absence of friction and stationarity of motion — the velocity of the geostrophic wind is perpendicular to the gradient, i.e., parallel to the isobar.

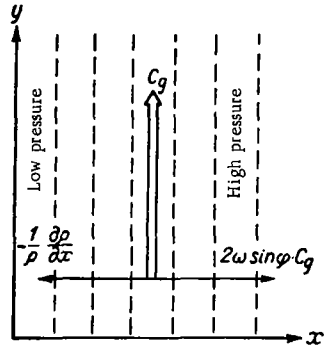


FIGURE 133. Geostrophic wind

Since the deflective force of the earth's rotation K acts to the right of the velocity in the northern hemisphere, it is obvious that the pressure force must act to the left of the velocity c_g. It also follows that the geostrophic wind blows along the isobars leaving low-pressure regions to the left and high-pressure regions to the right in the northern hemisphere; in the southern hemisphere the converse is true This relationship is called the barometric wind law

Figure 133 illustrates the directions of the geostrophic wind c_g and of the forces acting on a particle in this process; in accordance with what we said earlier, the pressure force G is balanced by the deflective force of the earth's rotation K, i.e.,

$$G + K = 0. \tag{30}$$

As (27) shows, the total speed of the geostrophic wind $c_g = \sqrt{u_g^2 + v_g^2}$ is given by the following formula:

$$c_g = \frac{1}{2\omega_z \rho} \frac{\partial p}{\partial n} = \frac{1}{2\omega \sin \varphi \rho} \frac{\partial p}{\partial n}. \tag{31}$$

Assuming $\rho = 0.001276\ \text{g/cm}^3$ (at $p = 1000$ mb and $t = 0°$) and $2\omega = 1.458 \times 10^{-4}\ \text{sec}^{-1}$ and expressing the pressure gradient in millibars per 111.2 km, we obtain

$$c_g = \frac{4.8}{\sin \varphi} \frac{\partial p}{\partial n}\ \text{m/sec}. \tag{31'}$$

Since the air density is not measured directly in meteorology, it is convenient to make use of the equation of state $p = \rho RT$ and transform (31) as follows:

$$c_g = \frac{RT}{2p\omega \sin \varphi} \frac{\partial p}{\partial n}. \tag{32}$$

This formula shows the dependence of the velocity of the geostrophic wind on temperature, pressure, pressure gradient and latitude. Thus for $\frac{\partial p}{\partial n} = 1$ mb/111 km and $\rho = 1.293\ \text{kg/m}^3$, the values of the geostrophic wind (in m/sec) are: 9.6 m/sec for $\varphi = 30°$, 6.8 m/sec for $\varphi = 45°$ and 5.5 m/sec for $\varphi = 60°$. The geostrophic wind also undergoes very substantial changes with the air density. For example, for the same pressure gradient and latitude, the velocity of the geostrophic wind in summer (when ρ is smaller) will be about 10% higher than in winter. Gradient rules, from which the geostrophic wind can be calculated very easily and rapidly, have been suggested for practical calculations by a number of authors.

At the beginning of the section we noted that the geostrophic wind approximates closely to the actual wind only under free-atmosphere conditions where the influence of friction can be disregarded. Near the surface, where the forces of friction are active, only some of the characteristics of geostrophic wind considered above are in in evidence. For instance, the velocity of the actual wind, though smaller than that of the geostrophic wind, is proportional, as is the latter, to the pressure gradient. Like the geostrophic wind, the real wind is deflected to the right with respect to the pressure gradient, not by 90° but (on the average) by about 60–70°; this is due to the influence of turbulent friction on atmospheric motions.

The difference between the actual and geostrophic wind is very significant in the low-latitude regions. For small φ the concept of geostrophic wind generally becomes meaningless, as is evident from formula (31), since for $\varphi \to 0$ the geostrophic velocity $c_g \to \infty$ (in this case the deflective force is zero).

b) Gradient wind for circular isobars. Let us now consider the stationary horizontal frictionless motion of air for the case of curvilinear isobars. In the general case of plane nonstationary curvilinear motion in

the absence of friction the equations of motion in Cartesian coordinates (25) have the following form:

$$\left.\begin{aligned}\frac{du}{dt} &= -\frac{1}{\rho}\frac{\partial p}{\partial x} + 2\omega_z v\\ \frac{dv}{dt} &= -\frac{1}{\rho}\frac{\partial p}{\partial y} - 2\omega_z u\end{aligned}\right\}. \tag{33}$$

We express these equations in polar coordinates (r, θ), by putting $x = r\cos\theta$ and $y = r\sin\theta$.

We denote by v_r and v_θ the components of the velocity in polar coordinates; then after simple transformations, instead of (33) we obtain the equations

$$\left.\begin{aligned}&\frac{\partial v_r}{dt} + v_r\frac{\partial v_r}{\partial r} + v_\theta\frac{\partial vr}{r\partial\theta} - \frac{v_\theta^2}{r} - 2\omega\sin\varphi v_\theta = -\frac{1}{\rho}\frac{\partial p}{\partial r},\\ &\frac{\partial v_\theta}{\partial t} + v_r\frac{\partial v_\theta}{\partial r} + v_\theta\frac{\partial v_\theta}{r\partial\theta} + \frac{v_r v_\theta}{r} + 2\omega\sin\varphi v_r = -\frac{1}{\rho}\frac{\partial p}{r\partial\theta}.\end{aligned}\right\} \tag{34}$$

Let us consider the particular case of plane frictionless stationary motion for concentric circular isobars. In this case

$$\frac{\partial v_r}{\partial t} = 0, \quad \frac{\partial v_\theta}{\partial t} = 0, \quad \frac{\partial p}{\partial\theta} = 0, \quad \frac{\partial v_\theta}{\partial\theta} = 0, \quad \frac{\partial v_r}{\partial r} = 0.$$

Owing to these conditions the second equation in (34) can be satisfied only for $v_r = 0$. From the first equation, replacing v_θ by c, we obtain

$$\frac{c^2}{r} + 2\omega\sin\varphi c = \frac{1}{\rho}\frac{\partial p}{\partial r}. \tag{35}$$

The above shows that the wind defined by this equation blows at right angles to the pressure gradient. Such a gradient wind with circular isobars is sometimes described as geocyclostrophic. Here the term $\frac{c^2}{r} = Z$ represents the centrifugal force.

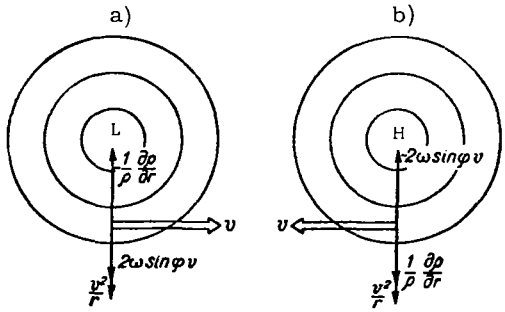


FIGURE 134. Gradient wind for circular isobars

a–in cyclone; b–in anticyclone.

Let us see how atmospheric motion for circular isobars in a cyclone (low-pressure region) and anticyclone (high-pressure region) is represented schematically. In the first instance the pressure force acts toward the center of the cyclone, and the centrifugal force Z and deflective force of the earth's rotation K, which balance it, act in the opposite direction,

as long as $\left|\frac{1}{\rho}\frac{\partial p}{\partial r}\right| > \left|\frac{c^2}{r}\right|$; when the latter condition is fulfilled the velocity will be directed counter-clockwise $\left(\frac{\partial \theta}{\partial t} > 0\right)$. Consequently, we have $G < 0$, $K > 0$ and $Z > 0$, and therefore $G = K + Z$. The relationship among the forces acting under these conditions is illustrated in Figure 134a.

In an anticyclone (Figure 134b) the pressure force G acts in the opposite direction and together with the centrifugal force Z balances the deflective force K, which acts in the direction of the center of the anticyclone. In this case equation (35) should be written as follows:

$$-\frac{c^2}{r} + 2\omega \sin \varphi c = \frac{1}{\rho}\frac{\partial p}{\partial r}, \tag{35'}$$

i. e., motion will be in the opposite direction— clockwise $\left(\frac{\partial \theta}{\partial t} < 0\right)$; further $G > 0$, $K < 0$ and $Z > 0$, and therefore $K = G + Z$, or $G = K - Z$. Thus in cyclones the wind blows in counter-clockwise direction and in anticyclones in clockwise direction.

In the general case the equation of the gradient wind for circular isobars can obviously be expressed as

$$\pm \frac{c^2}{r} + 2\omega \sin \varphi c = \frac{1}{\rho}\frac{\partial p}{\partial r}, \tag{36}$$

or

$$\pm Z + K = G, \tag{36'}$$

where the plus sign in front of the first term applies to cyclones and the minus sign to anticyclones.

For $r = \infty$ (straight isobars) equation (36) goes over into the equation of the geostrophic wind.

Let us rewrite (36) as follows:

$$c \pm \frac{c^2}{2r\omega_z} = \frac{1}{2\rho\omega_z}\frac{\partial p}{\partial r}.$$

The right-hand side of the above is c_g (cf. equation (31)). Consequently, in anticyclones, all other conditions being equal, the velocity of the gradient wind for circular isobars will be greater than the velocity of the geostrophic wind. In cyclones, on the contrary, $c < c_g$, i. e., cyclonic curvature of the isobars tends to reduce the wind speed by comparison with its geostrophic value. Thus for the same numerical value of the gradient, wind speed in an anticyclone will be greater than wind speed in a cyclone.

Looking at the relationship between the forces acting in anticyclones $(G = K - Z)$, we are led to the conclusion that in equatorial regions, where the deflective force of the earth's rotation $K = 2\omega \sin \varphi c$ is close to zero, the existence of stationary high-pressure regions must be impossible.

For a closer analysis of the question we rewrite equation (36) in a simpler form, introducing the notation $2\omega \sin \varphi = l$ and recalling that $\frac{1}{\rho}\frac{\partial p}{\partial r} = 2\omega \sin \varphi c_g = l c_g$:

$$\pm \frac{c^2}{r} + lc - lc_g = 0. \tag{37}$$

Solving this equation for anticyclones (with the minus sign in front

of $\frac{c^2}{r}$) and taking $c=c_1$, we have

$$c_1=\frac{rl}{2}\left[1\pm\sqrt{1-\frac{4c_g}{rl}}\right]. \tag{38}$$

In the above the minus sign in front of the square root should be retained, since for a positive sign we would arrive at the absurd conclusion that for r tending to infinity the velocity c also tends to infinity, whereas for $r=\infty$ it should be c_g. The maximum value of the velocity c_1 is restricted by the condition $1-\frac{4c_g}{rl}=0$ or $rl=4c_g$. But since $c_g=\frac{1}{2\rho\omega\sin\varphi}\frac{\partial p}{\partial r}$ while $2\omega\sin\varphi=l$,

$$\left(\frac{\partial p}{\partial r}\right)_{\lim}=\frac{1}{4}r\rho l^2=r\rho\omega^2\sin^2\varphi. \tag{39}$$

This shows that pressure gradients in anticyclones should be restricted by a certain limiting value $\left(\frac{\partial p}{\partial r}\right)_{\lim}$ given above, since for values of $\frac{\partial p}{\partial r}$ greater than this the expression under the radical in (39) would become negative.

Solving equation (39) for cyclones and denoting c by c_2, one has

$$c_2=-\frac{rl}{2}\left[1-\sqrt{1+\frac{4c_g}{rl}}\right]. \tag{40}$$

Arguing as before we arrive at the conclusion that pressure gradients and wind speeds in cyclones are in theory unbounded. This is supported by observations of tropical cyclones, in which very large gradients and wind speeds are recorded. However, under actual conditions in the atmosphere pressure gradients and wind speeds also have certain limiting values which are determined by the energy possibilities of the state of the atmosphere, and they cannot increase ad infinitum.

From the above statements it follows that, for the same gradient, the velocity of the gradient wind should be greater than the velocity of the geostrophic wind both in cyclones and in anticyclones. However, since under actual conditions cyclonic gradients are steeper than anticyclonic gradients, wind speeds are usually greater in cyclones than in anticyclones.

Chapter 22

WIND VARIATION WITH ALTITUDE

§ 1. Influence of friction on air motion

The relations obtained in the preceding chapter apply to cases where the forces of friction can be disregarded. In the lower atmospheric layers (to a height of 500–1000 m) this cannot be done and allowance must be made for the viscous forces. As a result of their action the wind observed near the ground differs from the geostrophic wind both in magnitude and in direction. For example, observations show that at a height of the order of 500 m the wind velocity is roughly twice as great as at the windvane height (about 10 m).

Furthermore, it has been established that, other conditions being equal, wind velocities over smooth even surfaces — and in particular over large expanses of water (e. g., seas) — are usually somewhat higher than over land, where departures from smoothness are present.

At the same time, winds near the earth's surface are deflected from the geostrophic direction toward regions of lower pressure by an angle of the order of 25–35°, which is smaller over smooth surfaces of water than overland. This angle of departure, as it is called, decreases with height, and with increasing altitude winds shift clockwise (to the right), reaching the direction of the geostrophic wind only at a certain height. It has also been established that the angle of departure at the surface depends on the curvature of the isobars: it is smaller in cyclones (of the order of 15–20°) and greater in anticyclones (30–35°).

Qualitatively all these departures of the actual wind from the geostrophic wind become understandable when one considers the influence of the viscous forces (forces of friction) on air motion. First of all we note that directly at the earth's surface, as near any fixed wall, the wind velocity always vanishes — the phenomenon called adhesion takes place. Starting from a certain height viscous forces due to atmospheric turbulence begin to play an important role: the result of their action is that the direction of the wind vector is always deflected away from the isobars towards the lower pressure and the wind direction begins to coincide with the isobars, as shown in the previous chapter, only from a certain height at which the influence of friction is negligible.

Before turning to a rigorous proof of the above let us see graphically what relationship should obtain between the forces acting upon a particle when the forces of friction $\vec{R}$ are taken into account. We recall that the pressure gradient $\vec{G}$ is always directed at right angles to the isobars towards decreasing pressure, while the deflective force of the earth's rotation $\vec{K}$ always acts at right angles to the direction of the wind velocity (to the right in the northern hemisphere). In the presence of the frictional

force $\vec{R}$, for steady motion where wind speed does not vary with time, the velocity $\vec{c}$ and all three forces $\vec{G}$, $\vec{K}$ and $\vec{R}$ should be in equilibrium.

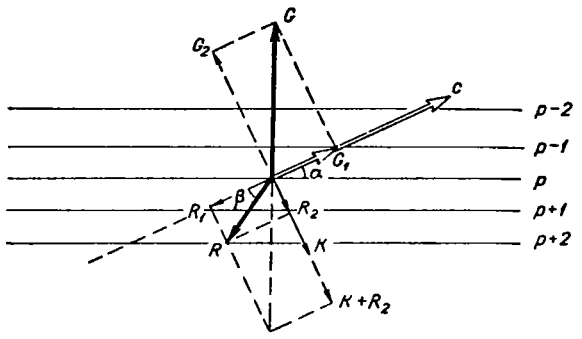


FIGURE 135. Stationary motion for rectilinear isobars in the presence of friction

It is easy to see that for rectilinear isobars stationary motion will be possible for the arrangement of the velocity $\vec{c}$ and forces $\vec{G}$, $\vec{K}$ and $\vec{R}$ shown in Figure 135, i.e., when the resultant of the forces $\vec{K}$ and $\vec{R}$ is equal in magnitude to $\vec{G}$ and opposite to it in direction. Indeed, if we consider the components of $\vec{G}$ and $\vec{R}$ along the direction of the wind and along the normal to it, for equilibrium the pressure froce $\vec{G}$ must be balanced by the resultant of the forces $\vec{K}$ and $\vec{R}$. But in this case

$$|G_1| = |R_1| \quad \text{and} \quad |G_2| = |K + R_2|,$$

or

$$|G\cos\psi| = |R\cos\beta| \quad \text{and} \quad |G\sin\psi| = |K + R\sin\beta|,$$

where ψ is the angle between the wind direction and the pressure gradient and β is the angle of deviation of the frictional force $\vec{R}$ from the direction opposite to that of the wind velocity $\vec{c}$; $\beta = \beta(z)$ varies with height.

According to these relations, equilibrium will be possible at any level only if the wind speed vector $\vec{c}$ is deflected from the isobar toward the lower pressure by the angle $\alpha = 90° - \psi$.

The distribution of all active forces for the case of circular isobars in which the centrifugal force $\vec{Z}$ comes into play is obtained in exactly the same way. Figure 136 shows that in a cyclone the resultant of the forces $\vec{K}$, $\vec{Z}$ and $\vec{R}$ balances the force $\vec{G}$, while in an anticyclone the resultant of $\vec{G}$ and $\vec{Z}$ is balanced by the resultant of $\vec{K}$ and $\vec{R}$. With this distribution of forces it is obvious that in cyclones the air near the surface flows towards the center from the periphery while in anticyclones, on the contrary, the air flows away from the center: in both cases motion takes place at a certain angle to the isobars.

The above clearly illustrates the barometric wind rule (Buys Ballot law) long known in meteorology· if an observer stands with his back to the wind, the low-pressure region will lie to his left and slightly ahead while the high-pressure region will lie to his right and slighly behind (in the northern hemisphere). This rule amends our earlier assertion

concerning the gradient wind (Chapter 21, § 4) according to which low pressure will lie to the left and high pressure to the right of the direction of motion. From the above discussion it is also seen that in cyclones the

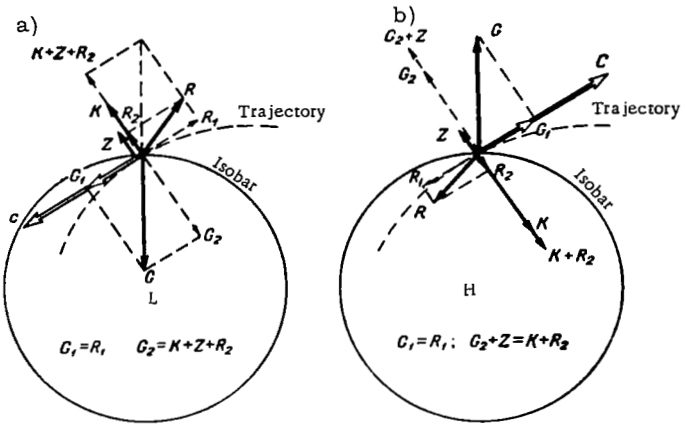


FIGURE 136. Stationary motion in the presence of friction in cyclone (a) and anticyclone (b)

counter-clockwise wind has a component towards the center as well, while in anticyclones it has a component away from the center.

§ 2. Wind profile in the surface layer

For a more rigorous treatment of the influence of friction it is necessary to turn to the complete equations of motion. In view of the fact that molecular viscosity (which is very small compared with eddy viscosity) may be disregarded, we write these equations, confining ourselves to horizontal motion, in the form:

$$\left.\begin{aligned}\frac{du}{dt} &= -\frac{1}{\rho}\frac{\partial p}{\partial x} + 2\omega_z v + \frac{\partial}{\partial z}\left(k_z \frac{\partial u}{\partial z}\right)\\ \frac{dv}{dt} &= -\frac{1}{\rho}\frac{\partial p}{\partial y} - 2\omega_z u + \frac{\partial}{\partial z}\left(k_z \frac{\partial v}{\partial z}\right)\end{aligned}\right\}. \qquad (1)$$

Since hereafter we will be dealing throughout with the coefficient of turbulence in the vertical direction, we agree for simplicity to drop the index z on k_z and write simply k.

It is perfectly obvious that the vertical distribution of wind speed is closely bound to the vertical variation of the coefficient of turbulence $k=k(z)$; we will therefore consider them together.

Turning to equations (1), we write them for steady motion $\left(\frac{du}{dt}=\frac{dv}{dt}=0\right)$:

$$\left.\begin{aligned}-\frac{1}{\rho}\frac{\partial p}{\partial x} + 2\omega_z v + \frac{d}{dz}\left(k\frac{du}{dz}\right) &= 0\\ -\frac{1}{\rho}\frac{\partial p}{\partial y} - 2\omega_z u + \frac{d}{dz}\left(k\frac{dv}{dz}\right) &= 0\end{aligned}\right\}. \qquad (2)$$

If the components of the geostrophic wind $u_g = \frac{1}{2\omega_z \rho}\frac{\partial p}{\partial y}$ and $v_g = \frac{1}{2\omega_z \rho}\frac{dp}{dx}$ are introduced, these equations become

$$\left.\begin{aligned} \frac{d}{dz}\left(k\frac{du}{dz}\right) &= -2\omega_z v + 2\omega_z v_g \\ \frac{d}{dz}\left(k\frac{dv}{dz}\right) &= 2\omega_z u - 2\omega_z u_g \end{aligned}\right\}. \tag{3}$$

We integrate the above equations over the altitude z; we then obtain

$$\left.\begin{aligned} k\frac{du}{dz} &= \left|k\frac{du}{dz}\right|_{z\to 0} - 2\omega_z\int_0^z (v - v_g)\,dz \\ k\frac{dv}{dz} &= \left|k\frac{dv}{dz}\right|_{z\to 0} + 2\omega_z\int_0^z (u - u_g)\,dz \end{aligned}\right\}. \tag{4}$$

Numerical evaluation of the terms in the right-hand side of the above shows that, within the surface layer (to a height of a few tens of meters), the second terms containing the small factor ω_z amount to no more than 10% of the first terms. For the surface layer these terms can be disregarded with sufficient accuracy for practical work and equation (4) can be written (approximately) in the following form:

$$\left.\begin{aligned} k\frac{du}{dz} &= \left|k\frac{du}{dz}\right|_{z\to 0} = \frac{\tau_{x,0}}{\rho} \\ k\frac{dv}{dz} &= \left|k\frac{dv}{dz}\right|_{z\to 0} = \frac{\tau_{y,0}}{\rho} \end{aligned}\right\} \tag{5}$$

where $\tau_{x,0} = \rho\left|k\frac{du}{dz}\right|_{z\to 0}$ and $\tau_{y,0} = \rho\left|k\frac{dv}{dz}\right|_{z\to 0}$ are the projections of the tangential stress of turbulent friction at the earth's surface (for $z=0$) on the x- and y-axis; these projections are sometimes called the surface friction.

According to (5), in the surface layer the turbulent stress, and together with it the wind velocity, remain constant in direction.

If the x-axis is oriented along the direction of the air flow, then $v=0$ $u=c$, $\tau_{y,0}=0$ and $\tau_{x,0}=\tau_0$, and instead of the system (5) we will have a single equation

$$k\frac{dc}{dz} = \frac{\tau_0}{\rho}, \tag{6}$$

where

$$\tau_0 = \rho\left|k\frac{dc}{dz}\right|_{z\to 0}. \tag{7}$$

Introducing the mixing length l one can write the coefficient of turbulence (as indicated earlier) as follows:

$$k = l^2\frac{dc}{dz}, \tag{8}$$

then, equation (6) assumes the form

$$l^2\left(\frac{dc}{dz}\right)^2 = \frac{\tau_0}{\rho} = v_*^2, \tag{9}$$

where $v_* = \sqrt{\frac{\tau_0}{\rho}}$ is the so-called dynamic velocity, or friction velocity.

But the mixing length l is a function of the height; moreover, as is the convention in aerodynamics, it can be taken as a liner function

$$l = \varkappa(z + z_0), \tag{10}$$

where $\varkappa = 0.38$ is von Kármán's constant and z_0 is the roughness length.

In view of (8) we transform equation (9) to

$$dc = \frac{v_*}{\varkappa} \frac{dz}{z + z_0}. \tag{11}$$

Integrating the above over the height from $z=0$, where $c=0$, to $z=z$, where $c=c(z)$, we obtain

$$c(z) = \frac{v_*}{\varkappa} \ln \frac{z + z_0}{z_0}. \tag{12}$$

Formula (12) is the law of variation of wind velocity in the surface layer of the atmosphere. It is known as the logarithmic law of vertical wind profile and was first obtained by Prandtl.

From the properties of the logarithmic function it follows that wind velocity rises rapidly with increasing height for small values of z, i.e., near the earth's surface; subsequently as z increases this rise becomes increasingly slow. Thus if the log z is plotted along the abscissa and the wind velocities c along the ordinate, according to our law (12) one should obtain a straight line.

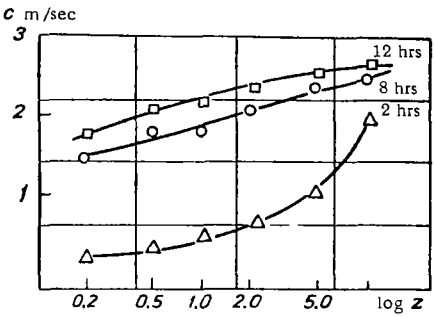


FIGURE 137. Observed vertical distribution of wind speed in the surface layer (Arys', Kazakh SSR)

However, analysis of a large amount of observational data shows that results calculated from formula (12) are in good agreement with experimental data only for a neutral thermal stratification of the surface layer, when $\gamma = \gamma_a$. For nonequilibrium thermal stratification, where $\gamma \gg \gamma_a$, or $\gamma \ll \gamma_a$, deviations of the wind profile from the logarithmic one are observed. The curvature of the wind profile in the semi-logarithmic coordinate system (c, $\ln z$) is positive for stable stratification ($\gamma < \gamma_a$) and negative for unstable stratification ($\gamma > \gamma_a$) of the surface layer. This is illustrated by Figure 137, which gives data of observations on 27 August at Arys' Station (Kazakh SSR) for stratifications of the surface layer close to unstable (12 hrs), neutral (8 hrs) and stable (2 hrs).

However, as was repeatedly mentioned earlier, cases of neutral stratification of the surface layer are rare. It was therefore very important to find a generalization of the above theory for the case of nonequilibrium temperature stratification.

The simplest way of constructing a scheme of the distribution of wind speed (as of other meteorological elements) in the surface layer is to approximate the wind profile by means of analytic formulas containing certain parameters which are then determined empirically as functions of the stability. The simplest scheme is that of the power profile according to which

$$c = c_1 \left(\frac{z}{z_1}\right)^n, \tag{13}$$

where c_1 is the wind speed at the initial level z_1 and n is a certain parameter which depends on the stability.

If this expression is inserted in the expression for the turbulent stress $\tau = \rho k \frac{\partial c}{\partial z}$, one finds that

$$k = k_1 \left(\frac{z}{z_1}\right)^{1-n}, \tag{14}$$

where $k_1 = \frac{\tau z_1}{\rho n c_1}$ is the coefficient of turbulence at the level z_1.

Obviously, n should be a proper fraction $(0 < n < 1)$. Its characteristic numerical values are roughly 1/7 for equilibrium conditions, 1/4 for inversion and 1/10 for super-adiabatic conditions. Experiments show that these relations turn out to be rather unsatisfactory. Whereas the actual wind profile can be expressed by formula (13) under inversion conditions, for equilibrium and super-adiabatic states there are considerable departures from it. Nonetheless these simple formulas are used to solve certain particular problems.

Schemes for the structure of the surface layer under nonequilibrium conditions proposed by M. I. Budyko and D. L. Laikhtman are more successful. Both authors start from the dependence of the mixing length l on the height z.

Budyko assumes that under nonequilibrium conditions the mixing length is also proportional to the height but with a different proportionality factor, i. e., the following dependence holds instead of (9)

$$l = \varkappa m z, \tag{15}$$

where the factor m characterizes the influence of the heat flux and is equal to unity in its absence.

Further, Budyko gives the dependence of the roughness length z_0' on the stratification in explicit form

$$z_0' = \frac{z_0}{m}, \tag{16}$$

which corresponds to vanishing of the average wind speed for the same value of the mixing length

$$l = \varkappa z_0. \tag{17}$$

It is easy to show that the formulas of Budyko's scheme are obtained from those given above by the simple substitution of $\varkappa m$ for $\varkappa$ and $\frac{z_0}{m}$ for z_0.

Experiments show that the quantity m varies roughly from 0.1 for deep inversions to 4 for very strong convection.

For the difference in wind speed at two levels we obtain

$$c - c_1 = \frac{v_*}{\varkappa m} \ln \frac{z}{z_1}, \tag{18}$$

and for the coefficient of turbulence

$$k = \varkappa m v_0 z, \tag{19}$$

so that according to this scheme the profile of k is linear.

In Laikhtman's scheme, which is somewhat different, the following relation is assumed for the mixing length under nonequilibrium conditions

$$l = Az^{1-\varepsilon}, \tag{20}$$

where ε is a parameter which depends on the stratification of the surface layer.

Making use of the above we obtain the following for the turbulent stress

$$\tau = \rho A^2 z^{2-2\varepsilon} \left(\frac{\partial c}{\partial z}\right)^2,$$

or

$$\frac{\tau}{\rho} = A^2 z^{2-2\varepsilon} \left(\frac{\partial c}{\partial z}\right)^2. \tag{21}$$

Taking the square root and integrating, we find

$$c = \frac{v_*}{A\varepsilon} z^{\varepsilon} + C. \tag{22}$$

If c_1 is the wind speed at the height z_1, so that

$$c_1 = \frac{v_*}{A\varepsilon} z_1^{\varepsilon} + c, \tag{23}$$

then, subtracting the above from (22), we have

$$c - c_1 = \frac{v_*}{A\varepsilon} \left(z^{\varepsilon} - z_1^{\varepsilon}\right). \tag{24}$$

Let $z_1 = z_0$ be the roughness level where $c_1 = 0$, then

$$c = \frac{v_*}{A\varepsilon} \left(z^{\varepsilon} - z_0^{\varepsilon}\right), \tag{25}$$

and for any given level

$$c_1 = \frac{v^*}{A\varepsilon} \left(z_1^{\varepsilon} - z_0^{\varepsilon}\right). \tag{26}$$

Dividing (25) by (26), we obtain the following for the wind profile instead of (25)

$$c = c_1 \frac{z^{\varepsilon} - z_0^{\varepsilon}}{z_1^{\varepsilon} - z_0^{\varepsilon}}. \tag{27}$$

The above is known as the generalized power law.

The parameter ε is zero under equilibrium conditions, positive $(0 < \varepsilon < 0.5)$ for inversions and negative $(-0.5 < \varepsilon < 0)$ for convective conditions.

Comparing the schemes of Laikhtman and Budyko one sees that the former gives a better description of the actual wind profile under inversion conditions.

For the profile of the coefficient of turbulence the above relations lead to the power law

$$\frac{k}{k_1} = \left(\frac{z}{z_1}\right)^{1-\varepsilon}, \tag{28}$$

according to which the coefficient of turbulence has a faster-than-linear increase with height under super-adiabatic conditions and a slower-than-linear one under inversions.

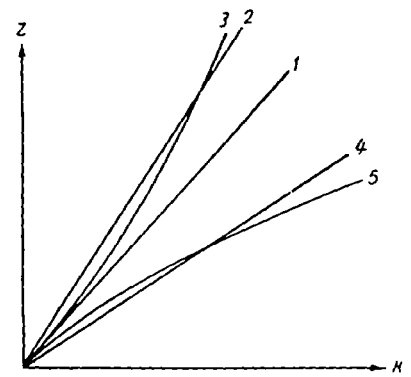


FIGURE 138. Profile of turbulence coefficient for different stratifications

1–equilibrium conditions; 2–inversion conditions (Budyko scheme); 3–inversion conditions (Laikhtman scheme); 4–convective conditions (Budyko scheme); 5–convective conditions (Laikhtman scheme).

Figure 138 gives the vertical variation of the coefficient of turbulence k in the surface layer for various temperature stratifications according to the schemes of Budyko and Laikhtman. Thus, the influence of stability on transfer is incorporated in different ways in the two schemes. However, both these schemes as well as the observational results indicate that the stratification of the surface layer exerts a very strong influence on the coefficient of turbulence. The attenuation of turbulence in inversion stratifications and its intensification for super-adiabatic vertical temperature gradients can change the coefficient of turbulence several times. The existing observational data provide the characteristic values of the coefficient of turbulence at a height of 1 m: of the order of 10^{-2} m^2/sec under medium inversion conditions, 0.1 m^2/sec under equilibrium conditions and 0.2–0.3 m^2/sec under medium convection conditions. Furthermore, for very strong instability the qualitatively new phenomenon of thermal convection of which we spoke in Chapter 6 comes into play.

In recent years the scheme of the structure of the surface layer under

nonequilibrium conditions has been considerably revised. A. S. Monin and A. M. Obukhov, applying the theory of similitude, showed that the velocity gradient in the surface layer should be represented in the form

$$\frac{\partial u}{\partial z} = \frac{v_*}{\varkappa z}\varphi\left(\frac{z}{L}\right),$$

where $L = \frac{-\rho c_p \overline{T} v_*^3}{\varkappa g P}$ is a characteristic length and P the turbulent flux of heat. The form of the function $\varphi\left(\frac{z}{L}\right)$ can be established for comparatively general assumptions.

Other schemes, on which we will not dwell here, have also been devised (L. T. Matveev, Priestley, et al.).

§ 3. Wind profile in the boundary layer of the atmosphere

In contradistinction to the surface layer, in which, as we have seen, the wind changes speed with height but remains practically constant in direction, at the higher levels of the atmospheric boundary layer the variation in wind speed is accompanied by a variation in direction. The nature of this variation, according to observational data, was briefly described in § 1.

In the lowermost few tens of meters one could assume with fair accuracy that turbulent friction does not change with height, which is equivalent to disregarding all forces other than eddy viscosity. But the influence of turbulence is also important in higher layers in which other forces come into play and must be considered— first and foremost the deflective force of the earth's rotation. To consider this question we will confine ourselves to a simple model in which the coefficient of exchange is constant with height.

Thus to simplify the problem we disregard the features of the surface layer (increase of k with height) and treat k as constant. Then, subject to the conditions that: the wind is horizontal, does not change in the k-direction and does not vary with time, the isobars are rectilinear; and, further, the pressure gradient does not vary with height, the system of equations of motion (1) become (directing the x-axis along the isobar):

$$\left.\begin{aligned} k\frac{d^2u}{dz^2} + 2\omega_z v &= 0 \\ k\frac{d^2v}{dz^2} - 2\omega_z u &= -2\omega_z u_g = \frac{1}{\rho}\frac{dp}{dy} \end{aligned}\right\}. \qquad (29)$$

As boundary conditions we specify that the wind speed vanishes at the earth's surface ($u_{z=0}=0$, $v_{z=0}=0$) and passes over into the geostrophic wind for an unlimited increase in height, i. e.,

$$u_\infty = u_g = -\frac{1}{2\rho\omega_z}\frac{dp}{dy} \quad \text{и} \quad v_\infty = 0.$$

Solving this system of equations we obtain the following expressions for u and v

$$\left.\begin{aligned} u &= u_g\left(1 - e^{-\zeta}\cos\zeta\right) \\ v &= v_g e^{-\zeta}\sin\zeta \end{aligned}\right\} \qquad (30)$$

where $\zeta = z\sqrt{\frac{\omega_z}{k}}$ is the dimensionless height.

The simplest way of arriving at this solution is to introduce the complex velocity

$$V = u + iv.$$

We multiply the second equation in (29) by i ($i=\sqrt{-1}$) and add it to the first. Recalling further that $v - iu = -i(u+iv) = -iV$ and that $-\frac{1}{\rho\omega_z}\frac{dp}{dy} = u_g$, we obtain

$$k\frac{d^2V}{dz^2} - 2\omega_z iV = -2\omega_z iu_g.$$

The general solution of this linear inhomogeneous equation is

$$V = Ae^{z\sqrt{2i\frac{\omega_z}{k}}} + Be^{-z\sqrt{2i\frac{\omega_z}{k}}} + u_g,$$

where A and B are complex integration constants.

Introducing the dimensionless height ζ, we write the general solution in the form

$$V = Ae^{\zeta\sqrt{2i}} + Be^{-\zeta\sqrt{2i}} + u_g.$$

The boundary conditions at infinity can be satisfied by setting $A=0$ in the above.

Consequently,

$$V = Be^{-\zeta\sqrt{2i}} + u_g.$$

In view of the fact that $B = B_1 + iB_2$ and $\sqrt{2i} = 1+i$, we have

$$V = u + iv = (B_1 + iB_2)e^{-\zeta(1+i)} + u_g.$$

Since from Euler's formula $e^{-\zeta i} = \cos\zeta - i\sin\zeta$,

$$V = (B_1 + iB_2)e^{-\zeta}(\cos\zeta - i\sin\zeta) + u_g =$$
$$= [B_1 e^{-\zeta}\cos\zeta + u_g + B_2 e^{-\zeta}\sin\zeta] + i[B_2 e^{-\zeta}\cos\zeta - B_1 e^{-\zeta}\sin\zeta].$$

From this it is evident that the conditions at the earth's surface ($z=0$ and, therefore, $\zeta=0$) can be satisfied by assuming $B_1 = -u_g$ and $B_2 = 0$, after which we finally obtain the equations (30) for the wind components.

For the angle of deviation of the wind from the isobar, i.e., the angle α between the wind direction and the geostrophic wind, we obtain from (30)

$$\alpha = \operatorname{arctg}\frac{v}{u} = \operatorname{arctg}\frac{e^{-\zeta}\sin\zeta}{1 - e^{-\zeta}\cos\zeta}, \qquad (31)$$

and for the wind speed (setting $u_g = c_g$)

$$c = \sqrt{u^2 + v^2} = c_g\sqrt{1 - 2e^{-\zeta}\cos\zeta + e^{-2\zeta}}. \qquad (32)$$

The above shows that wind speed increases with height from zero at the surface to a value equal to the velocity of the geostrophic wind c_g, which it assumes at the height given by the equation

$$2\cos\zeta = e^{-\zeta}, \qquad (33)$$

which gives $\zeta = z\sqrt{\frac{\omega_z}{k}} = 1.46$.

Expanding the indeterminate expression for $z=0$ in formula (31), we find that

$$\left|\lim \alpha\right|_{z\to 0} = \operatorname{arctg} 1 = \frac{\pi}{4}, \tag{34}$$

i. e., the wind deviation from the isobar at the ground is 45°; as the height increases α decreases and vanishes for $\zeta=\pi=3.14$, i. e., the height at which the wind coincides in direction with the geostrophic wind is considerably greater than the height at which its magnitude reaches the value c_g.

Plotting the velocity vectors at different heights in the coordinate system (u, v) and joining their ends, we obtain the hodograph of the dependence of u and v known as the Ekman spiral (Figure 139).

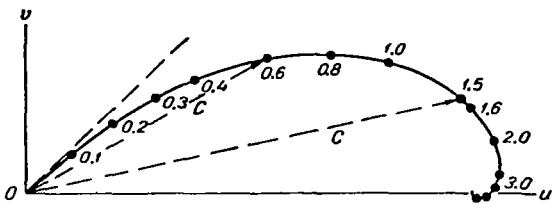


FIGURE 139. Ekman spiral

Several values of the height ζ are indicated on the figure near the curve; these include the value $\zeta = 1.6$, at which the speed of the geostrophic wind is reached, and $\zeta = 3.14$, at which the wind coincides in direction with the geostrophic wind.

Turning to the dimensional height z, which is related to ζ as

$$z=\zeta\sqrt{\frac{k}{\omega_z}}, \tag{35}$$

we arrive at the conclusion that the height at which the wind reaches the velocity and direction of the geostrophic wind increase as the square root of k and inversely as the square root of ω_z. Consequently, other conditions being equal, the wind will achieve the values of the gradient wind more rapidly in winter (small k) than in summer and more rapidly in high latitudes (large ω_z) than in low ones.

The above regularities apply to rectilinear isobars. Major mathematical difficulties arise for curved isobars, and even for simple circular isobars the formulas obtained are very complex. If one confines oneself to cases where the radius of curvature of the isobars is large compared with the height of the atmospheric layer under consideration, all the above formulas will remain valid; however, the dimensionless height ζ will be related to the dimensional height z as

$$z=\zeta\sqrt{\frac{k}{\omega_z+\frac{u_g}{r}}}, \tag{35'}$$

where r is the radius of curvature of the isobars, taken to be positive for cyclones and negative for anticyclones.

Qualitatively the above theoretical scheme gives a good explanation of the observed regularities in the vertical wind profile. From the quantitative point of view, however, it leads to values which differ substantially from the observed ones, the most significant departure being that the theoretical value of the angle of deviation (45° at the surface) is roughly twice the actual one ($\approx 20-30°$). This is a consequence of the fact that in deriving the formulas the coefficient of turbulence was assumed to be constant with height starting from the surface; in reality in the surface layer it increases linearly with height, and up to a height h in this layer the wind direction remains unchanged while the velocity varies logarithmically.

M. E. Shvets and M. I. Yudin allowed for this fact and obtained revised formulas. The results obtained from these agree entirely with reality. According to the revised theory, the speed varies only slightly, and only the value of the angle α changes; thus basically revision of the theory leads to modification of the Ekman spiral in the surface layer, and above this layer its character is not substantially changed.

§ 4. Vertical variation of wind speed and direction in the free atmosphere. Thermal wind

In the preceding section we considered the variation of the wind with height under the assumption that the horizontal pressure gradient is independent of height. Yet the horizontal pressure gradient does change with height both in magnitude and in direction. This variation is due to the temperature inhomogeneity of air masses in the horizontal direction, and it therefore follows that the vertical wind variation should be closely related to the presence of a horizontal temperature gradient in the atmosphere.

To grasp this point qualitatively, let us recall that in a warm column of air the pressure will decrease with height more slowly than in a cool column. As a result isobaric surfaces dip towards cooler masses. But then a pressure gradient inclined toward the cold mass will appear at any level; air motion which becomes geostrophic in the steady state (i. e., wind directed along the isobar) appears at this level under its influence.

Thus the presence of a horizontal temperature gradient modifies the pressure gradient and geostrophic wind which whould have prevailed in its absence.

To consider this we turn to the equations for the geostrophic wind, which we write in the form

$$u_g = -\frac{1}{2\rho\omega_z}\frac{\partial p}{\partial y}, \qquad v_g = \frac{1}{2\rho\omega_z}\frac{\partial p}{\partial x}, \tag{36}$$

or

$$c_g = \frac{1}{2\rho\omega_z}\frac{\partial p}{\partial n}.$$

From these equations it follows immediately that when the horizontal pressure gradient remains constant with height the velocity of the geostrophic wind increases with height owing to the decrease in the air density. To clarify the influence of the horizontal temperature gradient, consider a layer in the atmosphere situated between the levels z_0 and $z_0+\Delta z$, at which the temperatures are T_0 and T respectively. If the average

temperature of the layer is T_m, the pressure p at the upper level will be

$$p = p_0 e^{-\frac{g}{RT_m}\Delta z}, \tag{37}$$

where p_0 is the pressure at the lower level.

Taking the logarithmic derivatives with respect to x and y, we obtain

$$\left.\begin{aligned} \frac{1}{p}\frac{\partial p}{\partial x} &= \frac{1}{p_0}\frac{\partial p_0}{\partial x} + \frac{g\Delta z}{RT_m^2}\frac{\partial T_m}{\partial x} \\ \frac{1}{p}\frac{\partial p}{\partial y} &= \frac{1}{p_0}\frac{\partial p_0}{\partial y} + \frac{g\Delta z}{RT_m^2}\frac{\partial T_m}{\partial y} \end{aligned}\right\}. \tag{38}$$

Introducing $p = R\rho T$ and $p_0 = R\rho_0 T_0$ into the above and dividing the result by $2\omega_z$, we find

$$\left.\begin{aligned} \frac{1}{2\omega_z\rho}\frac{\partial p}{\partial x} &= \frac{T}{T_0}\,\frac{1}{2\omega_z\rho_0}\frac{\partial p_0}{\partial x} + \frac{gT\Delta z}{2\omega_z T_m^2}\frac{\partial T_m}{\partial x} \\ \frac{1}{2\omega_z\rho}\frac{\partial p}{\partial y} &= \frac{T}{T_0}\,\frac{1}{2\omega_z\rho_0}\frac{\partial p_0}{\partial y} + \frac{gT\Delta z}{2\omega_z T_m^2}\frac{\partial T_m}{\partial y} \end{aligned}\right\}. \tag{39}$$

Comparing the above equations with (35), one can write that

$$\left.\begin{aligned} (u_g)_z &= \frac{T}{T_0}(u_g)_{z_0} - \frac{gT\Delta z}{2\omega_z T_m^2}\frac{\partial T_m}{\partial y} \\ (v_g)_z &= \frac{T}{T_0}(v_g)_{z_0} + \frac{gT\Delta z}{2\omega_z T_m^2}\frac{\partial T_m}{\partial x} \end{aligned}\right\} \tag{40}$$

which shows that the components of the geostrophic wind at the level z are equal to its components at the level z_0 reduced by the factor $\frac{T}{T_0}$, plus certain additional terms called the components of the thermal wind. However, $\frac{T}{T_0}$ differs only slightly from unity; similarly $\frac{T}{T_m}$ is close to T, and therefore

$$\left.\begin{aligned} (u_g)_z - (u_g)_{z_0} &= -\frac{g}{2\omega_z}\frac{\Delta z}{T_m}\frac{\partial T_m}{\partial y} \\ (v_g)_z - (v_g)_{z_0} &= \frac{g}{2\omega_z}\frac{\Delta z}{T_m}\frac{\partial T_m}{\partial x} \end{aligned}\right\}. \tag{41}$$

From this it is evident that the differences in the components of the geostrophic wind between the two levels depend on the components of the gradient of average temperature in the layer between these levels:

$$\Delta c_g = \frac{g\Delta z}{2\omega_z T_m}\Gamma_m = \frac{g\Delta z}{2\omega \sin\varphi T_m}\Gamma_m = c_t, \tag{42}$$

where $\Delta c_g = c_t$ is the increment of the geostrophic wind in the layer Δz (from z_0 to z), in which the horizontal gradient of the average temperature T_m is Γ_m.

If the geostrophic wind at the initial level z_0 is $(\vec{c_g})_0$, then the wind at any other level z in the free atmosphere $(\vec{c_g})_z$ will be the vector sum

$$(\vec{c_g})_z = (\vec{c_g})_0 + (\Delta\vec{c_g}) = (\vec{c_g})_0 + \vec{c_t}. \tag{43}$$

Thus the thermal wind vector $\vec{c_t}$ is a supplementary term which, when added to the geostrophic wind at the lower level, gives the same at the upper level.

Expression (42) shows that the magnitude of the thermal wind is directly proportional to $\Delta z = z - z_0$ and to the horizontal gradient of the average temperature, and therefore the strongest winds are observed at greater altitudes where the horizontal temperature gradients are steepest.

Multiplying the first equation in (41) by $\frac{\partial T_m}{\partial x}$ and the second by $\frac{\partial T_m}{\partial y}$ and summing them, we obtain

$$[(u_g)_z - (u_g)_{z_0}] \frac{\partial T_m}{\partial x} + [(v_g)_z - (v_g)_{z_0}] \frac{\partial T_m}{\partial y} = 0; \qquad (44)$$

this shows that the thermal wind vector is normal to the horizontal gradient of the average temperature, i.e., lies along the average isotherms of the layer, and to the right of $\vec{\Gamma}_m$; that is, it stands in the same relation to $\vec{\Gamma}_m$ as the geostrophic wind to the pressure gradient $\vec{G}_{z=0}$. Thus the average horizontal temperature gradient plays the same role for the thermal wind as the pressure gradient does for the geostrophic wind.

The thermal wind can be calculated from formula (42): introducing the numerical values $g = 9.8\ \text{m/sec}^2$ and $\omega = 7.29 \cdot 10^{-5}\ \text{sec}^{-1}$ and expressing $\Delta z = z - z_0$ in meters and $\frac{\partial T_m}{\partial n} = \Gamma_m$ in degrees per 100 km, we find

$$c_t = 0.67 \frac{\Delta z}{T_m \sin \varphi} \Gamma_m \ \text{m/sec}. \qquad (45)$$

Let us apply these results to two arrangements of the isobars and isotherms which are of practical importance.

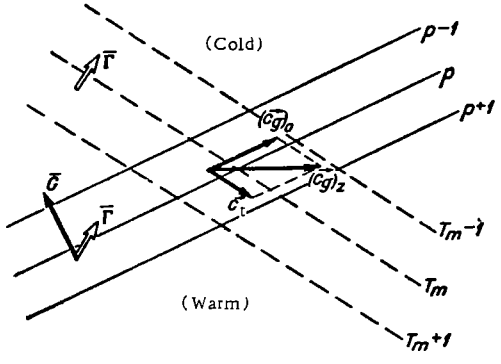


FIGURE 140. Right wind shift in the advection of heat

In the first case (cf. Figure 140) the horizontal temperature gradient $\vec{\Gamma}$ is directed to the right of the pressure gradient $\vec{G}$ at the initial level. General transport of air masses takes place from the warm region to the cold, and advance of air at a higher temperature (advection of heat) is observed at any point. The relationship between the vectors $(\vec{c_g})_z$ and $\vec{c_t}$ in Figure 140 was drawn in accordance with the above results and shows that the velocity vector $(\vec{c_g})_z$ shifts to the right with increasing height, approaching to the direction of the isotherms. Thus in a region of advection of heat the geostrophic wind shifts to the right with height.

In the second case (Figure 141) the horizontal temperature gradient $\vec{\Gamma}$ is directed to the left of the gradient $\vec{G}_0$. Advection of cold with general transport from the cold region to the warm region takes place. Figure 141 shows that in this case the velocity vector $(\vec{c}_g)_z$ will shift to the left as height increases, also seeking to approach the direction of the isotherms. It therefore follows that in a region of advection of cold the wind shifts to the left with height.

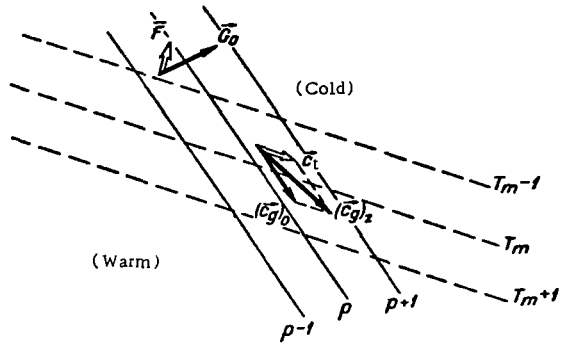


FIGURE 141. Left wind shift in the advection of cold

Similar reasoning can be applied to other mutual arrangements of the isotherms and isobars. For example, if the directions of the pressure and temperature gradients (G_0 and Γ_m), and along with them the isobars and isotherms coincide (which corresponds to the absence of advective heat transfer), we find that the wind becomes stronger with height without changing direction. If the temperature gradient Γ_m is opposite to the pressure gradient G_0 the geostrophic wind will slacken with height without changing direction, vanish at a certain height and subsequently assume the opposite direction if the original condition regarding the arrangement of the vectors $\vec{G}_0$ and $\vec{\Gamma}_m$ is conserved at all heights under consideration. The height at which this reversal in wind direction occurs depends on the value of the horizontal temperature gradient and wind speed at the lower level, i.e., on the value of the horizontal pressure gradient at the lower level.

These particular cases of wind change with height naturally do not exhaust the entire range of relationships actually observed.

Quite often right wind shift is observed in some atmospheric layers and left wind shift in other higher layers; this corresponds to advection of heat in the lower layer and advection of cold in the higher one. In such cases the stratification curve also varies with time— the atmosphere becomes more unstable.

Chapter 23

AIR MOTION IN THE BOUNDARY LAYER OF THE ATMOSPHERE

§ 1. Daily march of wind speed and direction

Observations show that in a steady weather regime a characteristic daily march of the wind velocity is clearly discernible in the lower layers of air.

Over land in the lowest layers (usually to a height of several meters) the maximum wind speeds are observed in daytime and the minimum wind speeds at night. In higher layers (to about one kilometer), on the contrary, maximum speeds are observed at night and minimum speeds by day. The amplitude of the daily march of wind speed amounts to 3–5 m/sec in middle latitudes. The vertical extension of the layers in which opposite daily marches are observed varies widely depending on the time of year. On hot summer days the upper boundary of the lower layer (so-called height of wind reversal) reaches up to 300 m; it drops to 20 m in winter inversions with mild wind. The upper limit of propagation of daily wind changes can vary between 500 and 2000 m.

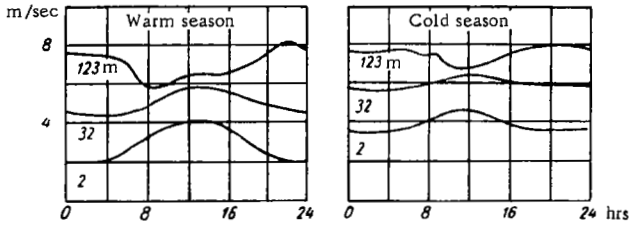


FIGURE 142. Daily march of wind speeds at various heights at Nauen

A sharp shift in air masses, the passage of fronts and other causes may, of course, disrupt the normal diurnal march of wind and lead to considerable wind changes from day to day.

A typical example of the daily march of wind speeds at different heights over land is given in Figure 142 (after observations at Nauen).

Analysis of observational data shows that, because of its smallness, the daily periodicity of the horizontal pressure gradient cannot produce significant changes in wind speed. Thus it is perfectly natural to link the daily wind march to another characteristic of the force field which determines air motion in the lower layers, namely the variability of the forces

of eddy viscosity, i. e., the daily variations of the vertical turbulent exchange. As we saw in Chapter 3, the coefficient of turbulence changes very considerably through the day, and the maxima of the exchange coefficient, which occur around noon, may be hundreds of times greater than the minima (at night).

The existence of a link between the daily march of wind speed and the intensity of vertical turbulent exchange is evidenced by an entire series of factors. The daily wind march is most clear-cut on clear, sunny days (with well-developed turbulent exchange) and is considerably fainter in overcast weather when differences in the intensity of exchange between night and day are greatly reduced and the magnitude of the exchange is small. Differences in the daily wind march between winter and summer correspond to differences in the daily variability of vertical exchange at these times of year. Diurnal variation of the turbulent exchange over the sea is usually small; this is accompanied by small daily fluctuations of wind speed.

Generally speaking, the influence of vertical exchange on the daily wind march reduces to the following. Owing to the daily periodicity of solar radiation at the earth's surface, in the surface layer of the atmosphere inversions are the rule at night and super-adiabatic gradients in daytime. In daytime, when thermal instability increases, turbulent exchange and, by the same token, interaction between upper and lower layers intensify. The more vigorous turbulence leads to a greater influx of slower particles from the lower levels to higher ones.

At hours with strong turbulent exchange one might therefore expect a reduction in wind speed in the upper layers (above 100–150 m) and increase in wind speed in the lower ones (to a height of 100–150 m). As exchange weakens the differentiation between layers begins to decrease towards evening and subsequently the opposite wind speed patterns become established in each layer; this is confirmed by observation.

This qualitative scheme cannot account for all details of the phenomenon; theoretical treatment of the question, taking into account the height-dependence of the coefficient of turbulence and the influence of many other factors, is necessary. However, the construction of a rigorous quantitative theory presents considerable difficulties.

The problem of the daily march of the wind speed and direction reduces to the solution of a system of equations in which the coefficient of turbulence is a function not only of the height but also of time. This nonstationary problem was solved by M. E. Shvets, A. T. Matveev, et al. for a specific daily march of the coefficient of turbulence, represented by a sinusoid in the simplest case. The form in which Shvets takes the coefficient of turbulence in the atmosphere (above the surface layer) is

$$k_1(t)\,h(t) = \overline{k_1 h}\,(1 + \varepsilon \sin \omega t).$$

Here $\overline{k_1 h}$ is the daily mean of $k_1 h$ and $\varepsilon \overline{k_1 h}$ is the amplitude of its daily march; as usual, $k_1(t)$ is the coefficient of turbulence at unit height and $h(t)$ is the height to which the increase in k with height z can, after the Shvets-Yudin model, be taken as linear and above which k remains invariant with height.

The theory developed by Shvets extends to any other form of the daily curve $k_1(t)$. Without going into the complex mathematical aspect of this theory, in which both k_1 and h are time-dependent parameters, we note that the formulas obtained permit fairly accurate computation of the daily

march of wind speed and direction. Concrete calculations with these formulas are in good agreement with direct measurements, especially for the upper layer. In particular they reveal that the height at which one march is replaced by the other— the so-called reversal height of the daily wind march— depends on $\overline{k_1 h}$ and thus on the time of year. In summer (for large $\overline{k_1 h}$) the reversal height is greater and amounts to hundreds of meters, while in winter (for small $\overline{k_1 h}$) it is smaller and amounts to tens of meters. From the theory, moreover, it follows that a daily march of wind speed is also observed above the boundary layer.

The direction of the wind also experiences regular changes during the day. For a steady weather regime, in the lower, boundary layer of air the wind shifts slightly to the right (in the direction of apparent motion of the sun) at the same time as it increases in speed from morning until 13–14 hrs; after midday it begins to shift to the opposite quarter, returning by evening to its initial direction. The reverse pattern obtains in the higher layers.

The cause which leads to these changes in direction is the same as that responsible for variations in wind speed— variations in the eddy viscosity.

In addition to the daily variations, an annual periodicity of wind speed and direction is also observed. This varies considerably from point to point and depends strongly on climatic features of the given area and local factors. To give an idea of the character of the differences observed we note that, for instance, near the western margins of continents in the middle latitudes of the northern hemisphere the maximum speeds are usually observed in winter and the minimum speeds in summer. At the same time inside vast land masses (e. g., Siberia) the lowest velocities occur in winter (calms are frequent) and the maximum wind speed is recorded in summer.

The annual march of the wind direction is even more varied and depends especially strongly on local conditions; however, we shall not consider this question here.

§ 2. Structure of the wind in the lower layers of the atmosphere

In the lower atmospheric layers the wind speed experiences disorderly, rapidly changing (in time) variations which superpose themselves on the principal (mean) velocity of transport of air masses. This phenomenon has a special name in meteorology— gustiness— and is, together with wind speed and direction, the object of painstaking study.

The sources of perturbation of the wind speed in the lower air layers are first and foremost the eddies formed in flows past various obstacles on the earth's surface. No clear ideas exist at present concerning the kinematic scheme of an eddy. As a very rough model eddies are sometimes pictured as closed circulation cells carried at the mean speed of the flow, something like a rotating cylinder with the axis in the horizontal plane. This model is supported to some extent by experimental data. Thus if one analyzes the synchronous recordings of two low-inertia recorders placed at an interval of 100 m along the direction of the wind, one will see that in many instances the rising and falling of wind registered by one instrument are shifted by an interval Δt from the time of registration of the same gusts by the other instrument. This interval is precisely the time required for

an eddy to move from one instrument to the other at the mean speed of flow. New eddies move with the air flow and penetrate into higher layers, gradually fading and breaking down into smaller eddy formations.

While the largest eddies are formed at the expense of the kinetic energy of general air-mass transport, smaller eddies derive their energy from larger eddies. The kinetic energy of very small eddies transforms directly into the thermal energy of molecular motion.

Thus, the atmosphere contains eddies belonging to a wide variety of scales and different stages of development. Studying the behavior of every individual eddy and constructing a kinematic model for it is therefore a very difficult task. In practice it is important to know the results of the interaction of an entire collection of eddy formations. Consequently, it is far more fruitful to study the general regularities of a set of eddies. Wind fluctuations are then treated as random perturbations and the characteristics of the latter are described by statistical methods.

The distribution of random velocity fluctuations in time and space, the correlations between individual components of the velocity, the dependence of the statistical characteristics of gusts on meteorological conditions (e.g., the temperature stratifications), relief, height above the ground and so forth— all these factors characterize the so-called wind structure.

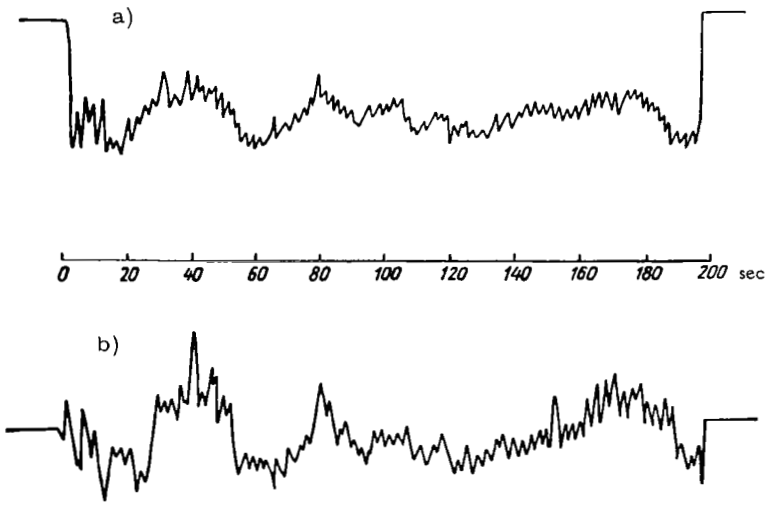


FIGURE 143. Character of wind fluctuations under convective conditions

a–horizontal component of wind speed; b–vertical component of wind speed.

The dimensions of eddies and the intensity of their penetration into overlying layers are largely conditioned by the temperature stratification. Under inversion conditions new perturbations are rapidly damped and the structure of the gusts is distinguished by relative uniformity, the mean amplitude and period of the fluctuations being limited.

Velocity records present an entirely different appearance under convective conditions (Figure 143): the uniformity of structure is disrupted, small gusts and large-scale perturbations exist side by side, the mean values of the amplitude and the periods of fluctuation increase. This is particularly noticeable on the record of the vertical component of wind speed. The reason is that under convective conditions the strong heating of the lowest air layers leads to the formation of updrafts which cause additional, purely thermal perturbations of the vertical component of velocity.

The question of the ratios between the different components of gustiness is of considerable interest. Specially designed experiments revealed that while the mean values of the fluctuations of the different velocity components were the same at heights of the order of 40 m, at a height of 1-2 m such isotropy did not occur: the velocity fluctuations in the horizontal plane were somewhat greater than the fluctuations of the vertical component.

According to A. R. Konstantinov, the ratio of the horizontal to the vertical components of gustiness under isothermal conditions amounts to 3.2 at the 2 m-level and drops to 1.3 at a height of 11.5 m. Analogous results were obtained in the experiments of other investigators.

This anisotropy of the wind structure in the lower layers is to be attributed to the influence of the earth's surface, which limits the development of vertical movements.

When the instability of the temperature stratification increases the vertical velocity fluctuations increase more rapidly than fluctuations in the horizontal plane and therefore anisotropy is somewhat reduced.

An important feature of the wind structure in the lower air layers is the presence of a correlation between the individual components of the velocity. According to experimental data, the coefficient of correlation between the vertical and horizontal components is always negative, i.e., larger vertical speeds correspond to lower horizontal speeds and vice versa. In absolute value this coefficient varies from 0.08 under inversion conditions to 0.35 for super-adiabatic gradients. When the coefficient of correlation and the vertical and horizontal components (according to wind direction) of gustiness are known, it is possible to determine such important characteristics of turbulence as the coefficient of exchange.

§ 3. Influence of obstacles on wind

The influence of local physico-geographic conditions results in certain areas in a number of characteristic features of the wind regime. Some of these can be treated as the effect of the earth's relief on general air currents. Indeed, as air moves over the earth's surface all unevennesses exert a mechanical influence on the flow, changing its direction and velocity and causing the appearance of a vertical component of velocity. This is due to the deformation and convergence or divergence of the streamlines which occur when the flow meets large and small landform features (mountain ranges, single mountains, hills, valleys, gorges, etc.)

Furthermore, uneven heating of the underlying surface due, for instance, to nonuniformity of the radiative characteristics of different areas (different albedo) or differences in their properties (land and water surfaces), or to differences in the inclination of exposed surfaces to the horizon, may

lead directly to the appearance of periodic winds in the lower atmospheric layers. Such winds are described as local.

Theoretical investigation of each of the many factors affecting the wind regime over dissected terrain and complex relief is possible only in particular cases and when the conditions of the problem are schematized. The study of wind features over dissected, nonuniform terrain thus reduces for the time being to obtaining experimental data and drawing qualitative inferences regarding the influence of given factors on the characteristics of wind.

The effect of large and small obstacles on wind is chiefly this: when an air stream meets a barrier, it flows past it. In the case of flow past a single hill, for instance, the advancing air usually flows around its flanks, changing direction chiefly in the horizontal plane; only when there is a strong wind will the air rise along the slopes and partly roll over the hill.

If air is moving between two barriers, e.g., hills, in the gap between them the streamlines will converge and the wind speed will rise. Due to this hollows, gorges and ravines have a considerable influence on the wind: the character of the wind in these cases will depend on the situation of these relief forms with respect to the direction of the principal air flow. If a hollow extends along the wind it will be well ventilated; if the wind is normal to the hollow air motion inside it will be feeble, since the greater part of the principal flow will pass over the hollow and not enter it. For the latter situation the temperature stratification has a very great influence on the character of the motion of the air filling the hollow.

Flow past hills is also accompanied by the creation of regions of intensive eddy formation on the leeward slopes (Figure 144). Owing to the law of

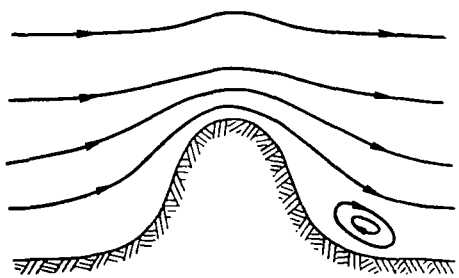


FIGURE 144. Wind flow past hill

conservation of mass, the velocity should be greater at the top of a hill than at its base (the streamlines converge at the top). As a result of this the pressure at the top is lower than that on the hillside (consequence of Bernoulli's law). Thus the pressure distribution of the windward side of the hill tends to accelerate the flow of air and that on the leeward side tends to retard it. Since air possesses viscosity, the motion of particles near the earth's surface is already substantially delayed. Due to this the pressure drop on the leeward side may not only halt particles moving in the lowest layer but may even cause motion in the opposite direction. This leads to the appearance of eddies which are subsequently removed from their place of origin and carried by the principal flow.

The theory of incompressible flow past obstacles is well developed in classical aerodynamics but cannot be applied directly to the conditions of the atmosphere since the latter is an inhomogeneous and compressible mediun in which eddy viscosity has an important influence.

The problem of flow past irregularities of the earth's surface was worked out with reference to the atmosphere by Academicians N. E. Kochin and A. A. Dorodnitsyn. Without going into the complex mathematical aspects of these researches we note that the principal conclusion to which they lead reduces to the following. The compressibility of air is an important factor in air flow past mountains; immediately above the mountain the streamlines rise with respect to the adjoining area, generally following the relief of the mountain, but at a certain large height above the mountain, on the contrary, the streamlines dip (Figure 145). This occurs over a fairly extensive layer of the atmosphere many times thicker than the height of the obstacle (even when the latter is low). At the same time so-called lee waves — undulating motions of air on the leeward slopes of mountains — are formed; pronounced updrafts often leading to cloud formation are observed in areas where such waves appear.

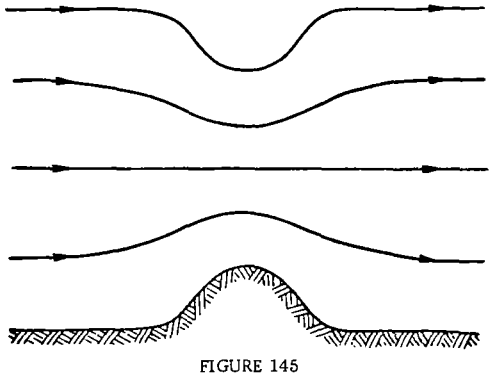


FIGURE 145

Special interest attaches to the question of the influence on the wind regime of shelterbelts planted in connection with drought prevention. Usually these consist of 10–20 m wide strips placed at right angles to the prevailing wind so as to form cells 200–500 m wide and over 1000 m long.

Shelterbelts protect fields by reducing the wind speed in the lowest layers of air. A kind of aerodynamic shadow forms behind the belt. This shadow gradually fills up as one moves away from the belt owing to descending motions and the disorderly penetration of eddies from the air thrown upwards at the edges of the belt. This reduction of the mean wind speed is also valuable in the prevention of dust storms, which remove fine soil particles together with humus and thus lower the fertility of the soil. Attenuation of the wind also reduces the transport of snow into hollows and gullies.

In addition, shelterbelts protect by modifying the size of the eddies contained in the advancing air flow. Tree trunks and branches act as an aerodynamic grid, breaking up the eddies which filter through them. The

reduction of the size of eddies beyond the belt lowers the intensity of turbulent exchange in the air layers close to the earth's surface; this reduces evaporation and significantly weakens the transport of snow and soil particles.

The aerodynamic influence of shelterbelts on wind, as that of any poorly streamlined obstacle, extends to the leeward and windward sides. The size of this zone of perturbation depends largely on the average height of the trees. Its downwind extension may be taken to be 30–40 such heights, and that in the opposite direction 10–12; for an average tree height of 10 m this gives 400 m (total size of perturbation zone). In practice appreciable perturbations (speed reduced by no less than 10%) extend over a horizontal distance 27–30 times greater than the height downwind and 1–3 times greater against the wind.

The size of the perturbation zone and value of the minimum average wind speed directly beyond the belt depend to a significant extent on the 'porosity' of the tree belt.

§ 4. Orographic winds. Foehn. Bora

The influence of orographic barriers on the wind regime is sharply manifested in a wide variety of forms in mountainous areas. Thus conditions which tend to strengthen wind are set up at passes and also in valleys and ravines.

Quite a few areas have special winds such as foehn and bora, as well as many related winds.

Foehn is a warm, dry wind blowing from mountains into valleys. In some cases it may cause air temperatures in a valley to rise by more than 10° in a few minutes; at the same time the relative humidity may drop by several tens of percents, frequently to values as low as a few percents.

Foehn is common to all mountain areas. It has been studied especially thoroughly in the Alps. In the USSR foehn is frequent in the Caucasus and also in Central Asia. It can last from a few hours to a few days at a time.

The number of days with foehn varies greatly with the features of the given mountain area. At Kutaisi the average number of days with foehn per year is 114 (38 in winter, 36 in spring, 13 in summer, 27 in autumn). In other parts of the Caucasus the number of days per year with foehn varies from a few to tens of days.

Conditions leading to the formation of foehn vary but typical situations are when an air mass travels across mountains, or when air moves down the leeward slope of an obstacle (mountain) in compensation for the subsidence of the air in the lower layers.

An elementary explanation for the temperature rise which occurs when air masses cross mountains is as follows. Air rises along the windward slope; this is accompanied by the condensation of water vapor and the release of the latent heat of vaporization, often with precipitation. At the same time, starting from the condensation level and up to the mountain peak, cooling of air on the windward side takes place according to the moist adiabatic, i.e., by less than 1° per 100 m of height. When the air descends on the other side of the mountain heating occurs, however, along the dry adiabatic, i.e., by 1° per 100 m of descent (Figure 146).

Precipitation on the windward side reduces the specific humidity of the air; therefore when the air descends along the leeward slope and is adiabatically warmed its relative humidity is significantly lower than before crossing the mountains. If the relative height of the mountain range is, say, 3 km and the mean moist adiabatic gradient is 0.6°/100 m, for an initial temperature of 20° at the bottom and condensation level of 500 m the temperature of the foehn flow at the foot of the mountain will be 30°, as shown by Figure 146. In reality the moist adiabatic temperature gradient is smaller at the lower level and greater at altitudes, and therefore the temperature of the rising air does not change uniformly; however, this does not alter the essence of the calculation.

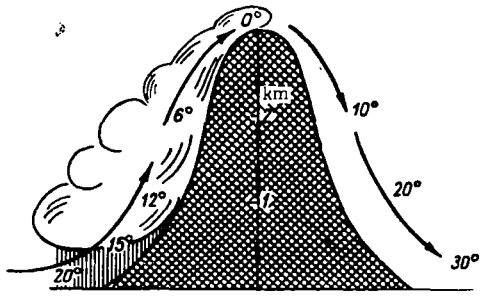


FIGURE 146. Foehn occurring as air mass crosses mountain

In cases where foehn is not directly related to an air mass travelling across the mountains, the temperature rise is determined by the value of the actual temperature gradient.

In the adiabatic descent of air the potential temperature remains constant; its variation with height is given by the formula

$$\frac{d\theta}{dz} = \frac{\theta}{T}(\gamma_a - \gamma).$$

The stabler the state of the air mass, the smaller the value of γ and the higher the potential temperature at altitudes by comparison with the surface layers.

Descending air in foehn carries a high potential temperature downwards, causing heating near the ground.

If $\gamma > \gamma_a$, i.e., the air mass is in unstable equilibrium, the potential temperature will decrease with height and upon development instead of a temperature rise there will be a certain temperature drop. Such examples of foehn are occasionally observed in the tropics.

From the very nature of the phenomenon it follows that foehn brings clear weather. Only at the mountain summits are clouds usually observed; these lenticular clouds remain immobile regardless of the strong wind that prevails there.

Prolonged warm, dry foehns are dangerous. Sharp temperature rises in the mountains lead to rapid melting of snow and overflowing of mountain streams. The very low humidity which accompanies the higher temperatures of foehn damages plants and can lead to crop failure over a wide area.

Although foehn is a local wind its appearance is possible only when definite weater conditions prevail in adjacent areas (e. g., cyclone in the foothills). In many cases this makes it possible to predict the onset of foehn.

Foehn and similar winds are known by various names in different regions. Examples are the "bokhorok"* of Sumatra and the chinook — strong foehn — on the eastern slopes of the Rocky Mountains. In the USSR similar winds are variously called: "ebe" (Dzhungaria and the Balkhash area), "kastek" (Kurdai pass between Alma Ata and Frunze), "ursat'evskii" wind (Fergana valley) "garmsil" (southern part of Central Asia), "afganets" (which, like the garmsil', produces violent dust storms), and many other apellations.

Bora is a strong gusty, cold wind which blows seaward down the slope of coastal uplands and causes considerable cooling in winter.

The Novorossiisk bora is particularly famous; when it is blowing the temperature in the bay may sometimes drop below −20° (or even lower) in wintertime, and wind speeds may exceed 60 m/sec in exceptional cases (at the pass). The number of days with bora per year at Novorossiisk is about 50 (bora is mostly observed in the November–March period).

The destructive force of the wind, freezing of water spray and icing of seaside structures and ships near shore sometimes have catastrophic consequences.

Let us consider the physical essence of the phenomenon. On the Black Sea coast near Novorossiisk the escarpment of the Varad ridge (with the Markhot pass) drops precipitously seaward. It is 400–650 m high and slopes gently towards the adjoining plain. When a powerful cold anticyclone with strong north-easterly winds developing along its southern edge occurs north of the area, the cold mass of the anticyclone gradually reaches the

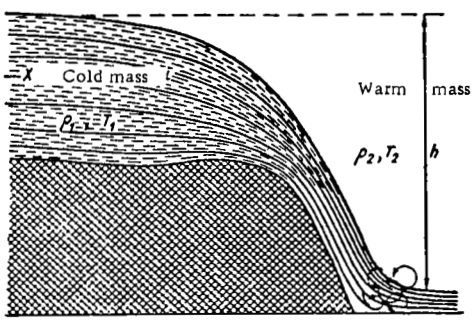


FIGURE 147. Bora

tall cliffs behind Novorossiisk and begins to tumble down (Figure 147). The rush of the cold, dense air down the pass (height 435 m) develops a considerable wind speed, as in the falling of water in waterfalls. Estimates show that the wind speed due to the descent of the cold air can reach values of the order of 25–30 m/sec. Adding to this the wind speed due to the horizontal pressure gradient and to the narrowing of the flow in the "neck" of the pass, we obtain the very high speeds observed during a bora.

The Novorossiisk bora is strongest in winter, as it is mainly at this time

* [Transliterated from Russian.]

that conditions for the incursion of very cold air masses into the area are favorable.

Over the sea the bora extends in all over a few kilometers (not more than ten). It is observed along the coast from Anapa to Tuapse and does not occur in other areas near Novorossiisk (beyond Tuapse).

The so-called "Baku nord" observed in the Baku area develops in the same way as bora. A typical bora is the north-westerly or north-north-westerly wind of the Baikal area ("sarma"), where a westerly wind analogous to "sarma" is also observed ("kharakanki"). Bora is also known in Novaya Zemlya and other places. In a Novaya Zemlya bora individual gusts can apparently reach 100 m/sec.

Aside from the winds considered here, mountain areas have many other wind forms including the winds in the gorges, passes and so forth mentioned above. An interesting wind observed in glacier-covered mountains, the glacier wind, as it is called, blows steadily down glaciers in the direction of their flow. As it moves it sucks air, as it were, from the surrounding mountainsides, so that in summer air temperatures over glaciers are always significantly higher than 0°; this promotes melting to no little degree.

The many other winds induced by orographic influences will not be described here.

§ 5. Local circulations. Breezes, mountain and valley winds

Whatever the scale of an air flow, its immediate cause lies in the non-uniform distribution of pressure in the horizontal plane. The appearance of horizontal pressure gradients can be due to thermal as well as dynamic factors. Indeed, it is well known that pressure drops less rapidly with height over warmer areas of the earth's surface, and isobaric surfaces in warm air are more widely spaced than those in cool air. At the same time isotherms are inclined to the horizon. This distribution of the isobaric surfaces can lead to the formation of a closed circulation of air.

Characteristic examples of such circulations are the b r e e z e s and the m o u n t a i n a n d v a l l e y w i n d s. We note that in this kind of motion, which usually covers limited areas, the air particles travel small distances over which the Coriolis force has only a slight effect. As a result the motion is determined chiefly by the horizontal pressure gradient and friction and takes place in a direction only slightly different from the gradient direction.

B r e e z e s are winds observed on sea and lake shores and are characterized by a regular reversal of direction during the day: at night they blow from land to water and in daytime, on the contrary, from water to land.

The first cause of coastal winds is the uneven heating of land and water. Overland in daytime intensive warming of the air leads to low pressure at the ground and high pressure at a certain height. Under such conditions (Figure 148) the pressure gradient and wind are directed towards the shore away from the water in the lower layer, and in the opposite direction at a certain height. This wnid is called the daytime or sea breeze. At night the directions of the pressure gradients and winds are reversed owing to the opposite temperature relationship.

Since the greatest contrast in temperature (and hence in air pressure) occurs in the immediate neighborhood of the shore, breezes reach maximum force near shore.

Depending on local conditions and weather, the sea breeze sets in between 8 and 14 hrs. Rising gradually, it reaches speeds of up to 6 m/sec near shore at the time of maximum development. It penetrates inland to distances of several tens of kilometers (20–30 and sometimes up to 100 km or even somewhat more). By evening the sea breeze drops and gives way to the converse flow, the nighttime land breeze, which rises gradually, at first, then falls and gives way to the day breeze after sunrise.

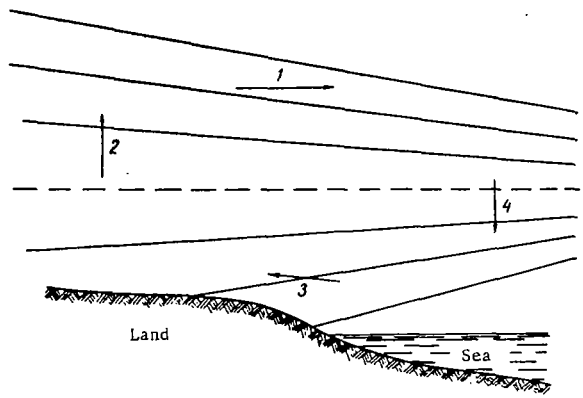


FIGURE 148. Schematic diagram showing distribution of isobaric surfaces in breeze circulation

The height to which the breezes extend can be as much as 1000 m; it is greater for the daytime than for the nighttime breeze, which usually ranges over a layer of about 100–200 m.

From the very essence of the phenomenon it follows that breezes are best developed in places where the daily temperature variation is largest and the sea-land temperature contrasts are maximum. Such conditions prevail in many tropical regions, where breezes are a year-round phenomenon. In middle latitudes breezes are far milder and are observed only in summer, since in winter the air over the coast is cooler than over the sea even in daytime. Throughout land breezes are milder with a smaller vertical and horizontal extension than sea breezes.

In the USSR the Black Sea and Caspian breezes are more vigorous than the Baltic breezes. At Sukhumi, for example, breezes are observed even in winter. Breezes are not felt on the northern seas.

Breezes also develop in a weaker form on the shores of comparatively small basins of water, such as Lakes Ladoga and Onega, the Tsimlyanskoe reservoir and other reservoirs.

Breezes and similar phenomena can appear only in weather with little cloud. Winds in cyclones often are many times stronger than the breezes, completely obscuring their action.

Another example of local circulation is the mountain and valley winds. These are periodic winds blowing down mountain slopes into valleys at night (mountain wind), and from valleys up mountainsides and the valleys themselves by day (valley wind).

Mountain and valley winds display a high degree of complexity in conformity with the diversity of conditions present in mountains (differences in the trend and depth of valleys, steepness of slopes, position of glaciers, etc.). We will therefore confine ourselves here to the general scheme of the phenomenon, disregarding the details.

Two types of mountain-valley circulations are distinguished: upslope and downslope winds, and proper mountain and valley winds.

The former constitute a circulation of limited vertical and horizontal extension and develop along the sides of valleys. In daytime air rises up the warm valley walls and is replaced at a certain height by air sinking down into the valley from higher layers. At night the pattern reverses itself: cool air flows downslope into the valley, and is replaced by warmer air masses coming from the atmospheric layers situated over the valley. This results in the rings of closed circulation illustrated schematically in Figure 149.

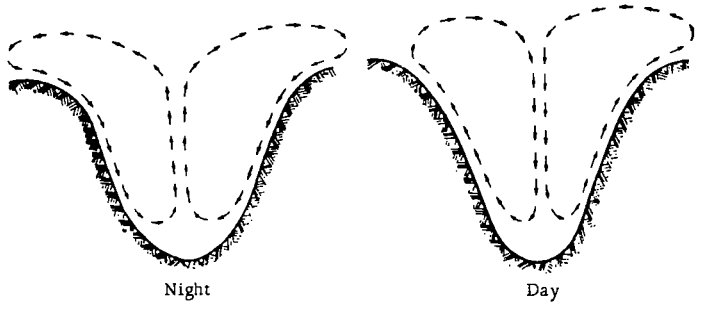


FIGURE 149. Mountain and valley winds (upslope and downslope)

Upslope and downslope winds usually appear only in summer and in clear weather. The qualitative interpretation of this phenomenon was given a long time ago and is evident from the above statements. The mathematical theory was developed recently (A. S. Monin and L. N. Gutman) and furnishes a good explanation of observed relationships.

The true mountain and valley winds are more complex. They blow along valleys, upwards in daytime (valley wind) and down the mountains at night (mountain wind). At a certain height (height of reversal) both winds change direction. The scale of this circulation is considerably greater (up to tens of kilometers) than that of upslope and downslope winds.

For illustration we present data obtained as a result of observations in the Caucasus by A. P. Khrgian (Tseiskoe gorge) and E. S. Selezneva (Baksan gorge). It was found that the more unstable the state of the atmosphere, the better developed the valley wind. It set in at about 7–8 hrs in the morning and reached a peak of about 2–4 m/sec on the average (maxima were up to 10 m/sec) after midday. In the evening (19–20 hrs) the valley wind dropped and was replaced at midnight by a mountain wind which

reached its peak of development before sunrise. It is interesting to note that while valley winds are observed only in clear weather, mountain winds are sometimes observed in overcast weather as well.

The height of wind shift to the opposite direction is a little more than 1 km (1.12 km) on the average in the Tseiskoe gorge and slightly lower in the Baksan gorge; for valley winds this height can be taken to be about 0.5–1 km above the valley bottom while for mountain winds it is appreciably lower (about 300 m). As to the vertical extension of the upper countercurrent, it can be traced all the way to the summits of the mountains surrounding the valley, above which it merges with the general currents of the free atmosphere.

The simplest explanation for the appearance of breeze and mountain-and-valley circulation is a qualitative one and is based on the so-called circulation theorem (cf. a textbook of fluid dynamics). As to the analytic solution of the problem of local circulations, this is an extremely difficult task. The mathematical difficulties involved stem from the fact that physically realistic schematization of the local winds is possible only if nonlinear convective terms are included in the equations of motion. However, no general methods for solving nonlinear problems are yet in existence. Nevertheless a number of particular solutions of the system of equations describing local circulations have been obtained by Soviet workers (I. A. Kibel', L. N. Gutman, Monin, Dorodnitsyn, et al.); these are set forth in detail in courses of dynamic meteorology.

Part Six
OPTICAL PHENOMENA IN THE ATMOSPHERE

In Part Two we considered direct and scattered radiation over the entire range of wavelengths, mainly with reference to questions of energy.

Radiation contains rays in the visible region of the spectrum which we perceive as light. When they travel through the atmosphere they produce a great many optical phenomena, the most important of which will be considered in the following chapters.

Chapter 24

OPTICAL PHENOMENA IN THE ATMOSPHERE DUE TO SCATTERING, ABSORPTION AND REFLECTION OF LIGHT

§ 1. Fundamentals of photometry with application to atmospheric optics

When studying radiant energy from the standpoint of the visual (light) sensations which it produces, it is usual to employ the terminology and units of photometry.

When the flux of radiant energy is evaluated on the basis of luminous sensations, it is referred to as the luminous flux, the unit of measurement of which is the lumen. Every receiver used for this purpose, and in particular the human eye, has a different sensitivity to different wavelengths Therefore the same radiant flux will evoke different sensations of brightness in different wavelengths. The ratio of the luminous flux to the radiant flux is called the luminous efficiency* or simply the luminosity. Figure 150

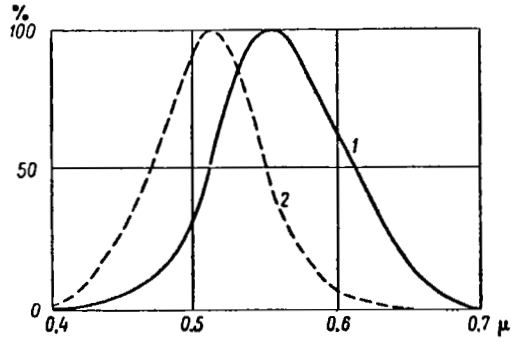


FIGURE 150. Relative luminosity curve for daytime (1) and twilight (2) vision

gives the curve of spectral sensitivity of the eye to monochromatic radiation λ, the so-called relative luminosity curve $v(\lambda)$.

* [Lit. "coefficient of luminosity" in the Russian.]

The solid curve in the figure refers to daytime conditions when the maximum sensitivity of the eye occurs at a wavelength of 555 μ (yellow-green), the dashed curve to twilight vision; the latter is altogether shifted toward the shorter wavelengths with a peak at $\lambda = 510\,m\mu$ (blue-green).

Denoting the monochromatic radiant flux by $d\Phi_\lambda$ and the luminous flux by dF_λ, we have

$$dF_\lambda = k_m v(\lambda)\, d\Phi_\lambda, \tag{1}$$

where k_m is a proportionality factor which depends on the choice of units for the two fluxes, the so-called luminous equivalent of power

In the case of the continuous spectrum of radiation the total luminous flux amounts to

$$F_\lambda = k_m \int_{\lambda=0}^{\lambda=\infty} v(\lambda)\, d\Phi_\lambda. \tag{2}$$

Numerically the luminous equivalent of power is $k_m = 680$ lumen/watt.

The fundamental photometric quantities are: luminous intensity, illuminance, luminous emittance and brightness.

The luminous intensity I of a point source is the luminous flux per unit solid angle, i.e., $I = \frac{d\Phi}{d\omega}$.

The unit of luminous intensity is the candle, defined according to an all-union standard kept at the All-Union Scientific Research Institute of Meteorology.

The illuminance of a surface is the density of luminous flux incident upon it, i.e., $E = \frac{dF}{ds}$. The unit of illuminance is the phot (one lumen per square centimeter), or the lux (equal to 10^{-4} phot).

If the luminous intensity of a point source is I then the illuminance of a surface normal to the rays at a distance r from the source of light will be $E = \frac{I}{r^2}$ and for arbitrary orientation of the surface $E = \frac{I \cos i}{r^2}$, where i is the angle of incidence of the rays.

The luminous emittance R is the ratio of the luminous flux emitted by a surface to its area, i.e., $R = \frac{dF}{ds}$. It is measured in the same units of flux per unit surface

The luminance B is measured by the luminous intensity of a surface in a given direction per unit area of the projection of this surface on a plane perpendicular to the direction under consideration, i.e., $B = \frac{I}{s \cos \varphi}$. The unit of luminance is the stilb.

If the luminous emittance gives the flux emitted in all directions, the luminance gives the flux emitted in a definite direction.

More often than not, in nature one deals with surfaces that send off reflected (diffuse) radiation rather than with self-luminous surfaces.

If a flux of density E falls upon an ideally matt surface of reflectance $A < 1$, its luminous emittance R will be $R = AE$, and since R, as we saw in Chapter 7, is related to the luminance B as $R = \pi B$, it is obvious that $B = \frac{AE}{\pi}$

If we express the luminance in stilbs then the luminous emittance of the surface will be π times greater and will be 3.14 lumen/cm^2. If the surface

is perfectly white, i. e., $A=1$, one can take the luminance of such a surface having a luminous emittance of 1 phot (i. e., emitting a flux of one lumen/cm^2) as the unit of luminance; this unit is called the lambert. A unit 10,000 times smaller is called the apostilb: this is the luminance of a surface which is emitting a flux of one lumen per square meter.

Obviously, one stilb = 3.14 lambert = 31,400 apostilb. However, it is far more convenient to replace this usual photometric definition of luminance by one which is essentially identical to it, namely: the luminance B of an object along a given direction is the ratio of the illuminance E produced by it on a surface element placed at right angles to the direction to the object, to the solid angle ω subtended by the object at the surface element, i. e., $B = \frac{E}{\omega}$.

Very broad ranges of illuminance (from 0.0002 to 10^5 lux) and luminance (from $5 \cdot 10^{-10}$ to 150,000 stilb) are encountered in nature. The eye possesses the ability to adjust itself to existing features and modify its sensitivity accordingly. This process of adjustment, as is well known, is called the adaptation of the eye. However, the sensitivity of the eye is not unbounded and it ceases to perceive when the quantity of light entering it falls below a minimum; this minimum is called the luminous threshold of vision.

In the visual perception of luminous fluxes the eye is sensitive not only to brightness but also to color, which is why nature is rich in color during daytime.

Aside from the group of achromatic colors, including white, black and intermediate tones of gray, all the remaining, so-called chromatic, colors differ from each other in brightness, hue and saturation.

Table 69 lists wavelengths corresponding to different colors.

TABLE 69

Color	Interval λ (mμ)	Typical λ (mμ)
Violet	390–455	430
Cyan	455–485	470
Blue	485–505	495
Green	505–550	530
Yellow-green	550–575	560
Yellow	575–585	580
Orange	585–620	600
Red	620–760	640

Many of the colors can be obtained by mixing white with a certain monochromatic spectral color. This allows one to characterize any color by means of two basic quantities: 1) the hue, which is the quality of the color expressed by the wavelength λ of the monochromatic light added to white, and 2) the saturation, or purity, which determines the proportion of this monochromatic light in the mixture; thus the saturation is given by $p = \frac{B_\lambda}{B_\lambda + B_w}$, where B_λ and B_w are the brightness of monochromatic and white color, respectively Obviously, for white $p=0$ and for a pure spectral color $p=1$.

Colors can also be characterized more precisely by describing all the possible hues in terms of three basic colors mixed in varying proportions. Then, any color will be obtained by adding a certain quantity of white to this mixture. Red, green and cyan have been chosen as basic colors.

§ 2. Apparent form of the sky and related phenomena

Observations indicate that the sky always seems to us to be flattened vertically and that we do not see it as a hemisphere overhead. This is to be attributed to psychological factors and also in part to physiological peculiarities of visual perception which need not detain us here.

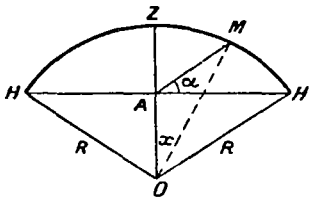


FIGURE 151

Suppose HH in Figure 151 is the horizon and an observer is standing at the point A; then, if we bisect the celestial arc ZH between the zenith and the horizon visually, we find that the angle α will, on being measured by any suitable instrument, prove to be roughly 22°, and not 45° as would be the case if the arc HZH were a semicircle. This visually determined angle α can be taken as the measure of the oblateness of the sky. After determining it one can also find the apparent distance to the zenith ZA and to the horizon AH, and also their ratio.

Indeed, as is easily seen from Figure 151, if we take the arc HZH as the arc of a circle of radius R with center at the point O we will have

$$\left.\begin{aligned} AZ &= R - OA = R - R\cos 2x = 2R\sin^2 x \\ AH &= R\sin 2x \end{aligned}\right\} \qquad (3)$$

where x is the angle between the directions OM and OZ; it is easy to show that it is related to α by

$$\operatorname{tg}\alpha = \frac{\cos x - \cos 2x}{\sin x}. \qquad (4)$$

The ratio $AH : AZ = \operatorname{ctg} x$, which characterizes the oblateness of the sky within the range of observed values of α, will be:

α (°)	18	20	22	24	26	28	30	32
$AH : AZ$	4.49	3.98	3.55	3.19	2.88	2.60	2.36	2.14

Observations show that the oblateness of the sky does not remain constant. It is greater (α is smaller) for strong illumination of the sky, i. e., in daytime, than at night; it also tends to be greater in the presence of clouds, smoke, etc.

Owing to this apparent oblateness of the sky, the altitude of points sighted in it is always estimated erroneously in visual observations. Thus the actual heights are always smaller than those estimated visually: at the same time, the same arc will be overestimated to heights of about 35° and underestimated at greater heights, Near the horizon, for example, an arc of 5° will be overestimated by a factor of nearly two, while near the zenith, on the contrary, it will be underestimated by a factor of two (Table 70).

TABLE 70

True elevation (°)	Visual estimates of elevation				True elevation (°)	Visual estimates of elevation			
	daytime ($\alpha = 22°$)		night ($\alpha = 30°$)			daytime ($\alpha = 22°$)		night ($\alpha = 30°$)	
	elevation	difference	elevation	difference		elevation	difference	elevation	difference
5	13.3°	11.4°	9.6°	8.6°	50	70.7°	3.0°	63.7°	3.9°
10	24.7	9.4	18.2	7.7	55	73.7	2.7	67.6	3.6
15	34.1	8.1	25.9	7.1	60	76.4	2.5	71.2	3.4
20	42.2	6.7	33.0	6.4	65	78.9	2.4	74.6	3.3
25	48.9	5.7	39.4	5.8	70	81.3	2.3	77.9	3.1
30	54.6	4.9	45.2	5.2	75	83.6	2.2	81.0	3.0
35	59.5	4.2	50.4	4.9	80	85.8	2.1	84.0	3.0
40	63.7	3.7	55.3	4.4	85	87.9		87.0	
45	67.4	3.3	59.7	4.0					

This gives rise to a whole series of phenomena. Thus when observing the sun and moon we overestimate their angular dimensions if they occur near the horizon and underestimate them at elevations greater than 35°. When they rise and set the sun and moon appear 4–5 times larger than they do at an elevation of about 60° For the same reason clouds occurring close to the horizon always appear to occupy a larger portion of the sky than is really the case. Finally, we usually overestimate considerably the height of objects protruding beyong the line of the horizon (mountains, towers, etc.)

The above considerations should always be borne in mind when estimating the dimensions of phenomena observed in the sky.

§ 3. Luminance of the sky

As a result of the scattering of the direct solar rays in the atmosphere every element of volume of the latter becomes a source of scattered light and the sky possesses a definite luminance in any direction. For a theoretical calculation of the luminance of a cloudless sky in the first approximation, one can take Rayleigh's theory as point of departure and make use of the considerations of Chapters 8 and 9. In Chapter 9 we derived a formula for the intensity of monochromatic scattered radiation, which in photometric terms is simply the luminance (19). Denoting the luminance

by $B_{\lambda,\varphi}$, we can write this expression as

$$B_{\lambda,\varphi} = \frac{3}{16\pi} S_{0,\lambda} v(\lambda) \frac{m}{m_\odot - m} \left[e^{-\sigma\lambda M_0 m} - e^{-\sigma\lambda M_0 m_\odot}\right](1 + \cos^2\varphi), \qquad (5)$$

where all symbols are as previously and $v(\lambda)$ is relative luminosity.

Integration of the above over the wavelengths of the visible spectrum gives the photometric luminance of the sky in the direction of the scattering angle φ.

We recall, however, that this formula was obtained without considering absorption, scattering of second and higher orders, polarization and diffuse reflection by the earth's surface. These factors can be calculated on the basis of Chandrasekhar's theory, developed with reference to the solution of the problem of Rayleigh scattering in planetary atmospheres. An exposition of this complex theory would take us too far afield.

For illustration results of calculation of the sky's luminance in the sun vertical are presented in Table 71, and the angular distribution of luminance in the sun vertical as calculated from Chandrasekhar's theory (after Sekera) in Figure 152.

However, all theoretical calculations refer to an ideal (molecular) atmosphere and give only a general schematic picture, neglecting the influence of light scattering on the large aerosol particles always present in the atmosphere. It is therefore of particular interest to consider data of direct observations and compare them with the theoretical results.

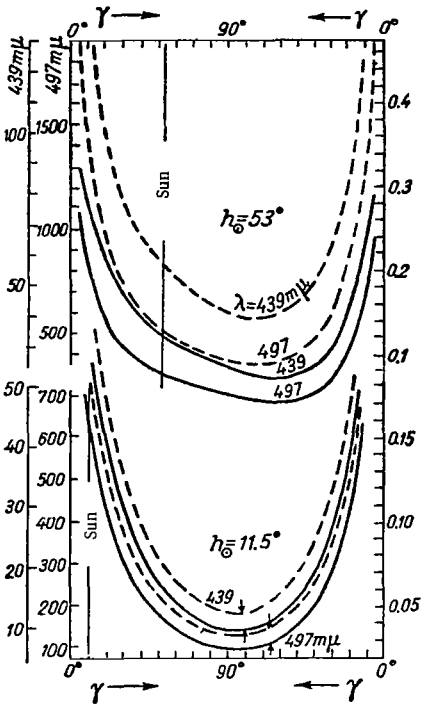


FIGURE 152. Angular distribution of luminance in the sun vertical as calculated from Chandrasekhar's theory (after Sekera); wavelengths λ= 439 and 497 mμ, $\Delta\lambda$ = 10 mμ; solar elevations $h_\odot = 53°$ ($z_\odot = 37°$) and $h_\odot = 11.5°$ ($z_\odot = 78.5°$).

It is convenient to express results of observations of the distribution of luminance over the sky by means of isophots — lines of equal luminance in the sky. Maps of isophots compiled by E. V. Pyaskovskaya-Fesenkova are shown in Figure 153. The maps were drawn for different $z_\odot$ and the same transmission coefficient (p= 0.80); the numbers near the curves indicate the luminance in stilbs.

When experimental and theoretical data are compared it is found that they agree The following fundamental regularities may be formulated for the distribution of luminance over a cloudless sky: 1) the half of the sky containing the sun is brighter than the other half; 2) regions close to the horizon are brighter; 3) there is a strong intensification of the brightness around the sun (so-called solar aureole); 4) as the sun's elevation

decreases there is a reduction in the luminance of the sky, especially in the half opposite the sun; 5) there is a tendency to form a luminance

TABLE 71

Elevation of sky in sun vertical for $z_\odot = 60°$ according to Rayleigh's theory

	Elevation of point in sky (°)							
	90	85	60	0	30	60	85	90
	Towards sun				Away from sun			
Luminance*(10^3 apostilb)	39.6 42.9	26.6 28.6	4.2 4.9	2.6 3.2	2.4 3.0	5.2 6.0	24.4 26.4	39.6 42.9
Hue λ ($m\mu$)	562	485	478	478	478	479	486	562
Saturation (%)	7	17	36	38	37	34	18	7

* Allowance for single scattering in first row, multiple scattering in second row.

minimum at a distance of about 90° from the sun; 6) light reflected by the earth s surface has an influence (when the albedo increases the luminance of the sky also increases).

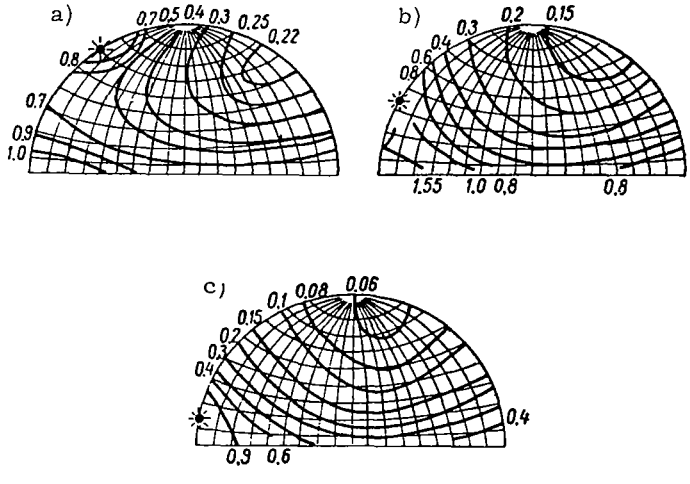


FIGURE 153. Luminance of sky in stilbs for $p = 0.80$

a- $z_\odot = 30°$; b- $z_\odot = 60°$; c- $z_\odot = 80°$.

§ 4. Blue color of the sky

A cloudless sky will display several distinct tones of blue in different parts An early attempt to explain this was made by Leonardo da Vinci but the correct explanation was discovered only after the theory of light scattering in a turbid medium had been developed.

Since the (molecular) scattering of light takes place in inverse proportion

to λ^4, in the spectrum of scattered light from the sky the maximum of the energy distribution should be shifted towards the short wavelengths and the light perceived from the sky should be blue. This is confirmed by calculations, as may be seen from the last two rows in Table 71, which gives the hues and saturations of the scattered light from various points of the sun vertical.

These data show that hue is characterized by values of the order of $\lambda = 475-490\,m\mu$ and shifts to $\lambda = 560\,m\mu$ only at the horizon, where, moreover, the saturation, i.e., blueness, is smaller; the point of highest saturation lies at a distance of roughly 90° from the sun.

In the actual atmosphere one must consider the influence of absorption and especially of scattering on large particles. Owing to this influence in the turbid atmosphere scattering depends to a lesser degree on wavelength, and the saturation of the blue color of the scattered light is considerably lower, than in the Rayleigh (dry and pure) atmosphere.

These theoretical considerations are supported by observational data. Table 72 gives the observed spectral composition of scattered light from a clear sky and a sky with cloud (in relative units). One can see from the table how much larger the range of short wavelengths in a clear sky is in comparison with a sky covered by clouds, or with direct solar light.

TABLE 72

Spectral composition of scattered light (relative units)

$\lambda\,(m\mu)$	Clear sky	Light cloud veil	White clouds	Sun
670	0.30	0.83	0.46	0.39
630	0.53	0.57	0.69	0.62
590	1.00	1.00	1.00	1.00
550	2.31	2.24	2.14	1.87
510	5.75	4.82	4.30	3.63
470	18.17	13.34	11.87	8.79
430	61.63	36.52	30.73	19.74

Observations show that the blueness of the sky experiences both diurnal and yearly fluctuations; it generally decreases with increasing $z_\odot$. Further, a close correlation with the turbidity (transparency) of the atmosphere has been established: the greater the turbidity, the less saturated the blue and the more whitish the sky. On the basis of this correlation certain authors point to a connection with the character of air masses. The blueness is also different in different geographic conditions. The saturation is greatest in the pure air of mountain regions and steppes, smaller in mid-ocean and on sea coasts, and particularly low in industrial zones (cities) with their dusty atmosphere.

With increasing elevation above sea level the color of the sky becomes increasingly cyan-blue and at 15–20 km it begins to take on a violet tone. Above 300 km, according to Cosmonaut Yu. A. Gagarin, the sky is completely black.

It is common knowledge that an overcast sky is gray throughout. The admittedly meager observations available indicate that the luminance of a sky covered by an unbroken layer of low cloud decreases from zenith to

horizon by roughly 2.5–3 times; at the zenith, moreover, the clouds seem somewhat yellowish while a bluish tinge appears at the horizon. In many cases darker patches, occasionally slightly bluish in tone, appear against the background of the cloud ceiling; this is to be attributed to differences in reflectance between different sections of the earth's surface.

§ 5. Polarization of light scattered by the atmosphere

The theory of light scattering shows that scattered light must be polarized. According to Rayleigh s theory, the degree of polarization, which is expressed by the parameter $p=\frac{i_\perp-i_\parallel}{i_\perp+i_\parallel}$, will be

$$p=\frac{i_\perp-i_\parallel}{i_\perp+i_\parallel}=\frac{1-\cos^2\varphi}{1+\cos^2\varphi}=\frac{\sin^2\varphi}{1+\cos^2\varphi} \qquad (6)$$

Thus according to Rayleigh's theory light should be polarized in the plane perpendicular to the plane of vision. Further, complete polarization ($p=1$) will be observed in directions perpendicular to the direction of the incident light ($\varphi=90$ and 270°); in directions coinciding with and opposite to the incident light ($\varphi=0$ and 180°) polarization will be zero ($p=0$), i. e., we will have unpolarized light. This is shown by the polar graph given in Figure 45 (Chapter 8), which we supplement by presenting the distribution of the degree of polarization as a function of the scattering angle φ (Figure 154).

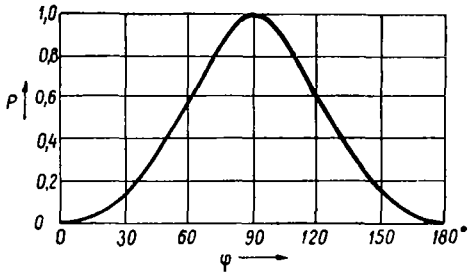


FIGURE 154. Degree of polarization according to Rayleigh's theory

In this case the plane of polarization would coincide with the plane passing through the sun, the sighted point and the eye of the observer. However, this situation does not correspond to observations and complete polarization is never observed in the actual atmosphere; the peak of polarization reaches values of about 85% only occasionally. One of the reasons for incomplete polarization, as we saw above, is that scattering particles are optically anisotropic. If this circumstance is taken into account in place of (6) we obtain

$$p=\frac{(1-\rho)\sin^2\varphi}{1+\cos^2\varphi+\rho\sin^2\varphi}, \qquad (7)$$

where ρ is the so-called depolarization factor, equal to 0.043 for air.

From (7) we obtain $p_{max}=0.922$ (instead of 1) for the maximum possible polarization.

In addition it is surely necessary to allow for the effect of higher-order scattering, and also for the fact that scattering on large particles, which does not give complete polarization and is considerably smaller than molecular scattering, also takes place in the atmosphere.

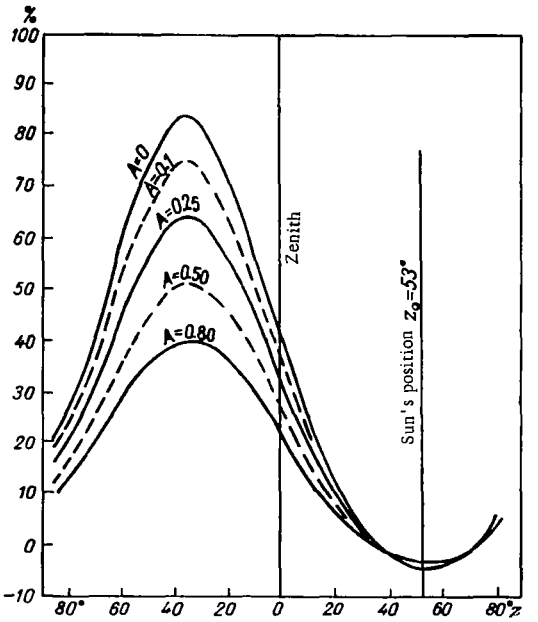


FIGURE 155. Distribution of the degree of polarization in the sun vertical for different albedos

The effect of higher-order scattering is primarily that on multiple scattering light is polarized in a direction other than that prevailing in single scattering. This reduces the degree of observable polarization, excepting those places where the degree of polarization of singly-scattered light is roughly zero, and also alters the position of the plane of polarization. Multiple scattering is strongly felt, for example, in the case of large surface albedos; it is particularly important for rays of small wavelength. These effects can be dealt with on the basis of the above-mentioned theory of Chandrasekhar, bringing the calculated results closer to the observed facts As illustration we give the distribution of the degree of polarization along the sun vertical for different albedos, as calculated from Chandrasekhar's theory (Sekera), in Figure 155.

The polarization of the light scattered by the atmosphere was discovered at the beginning of the last century It was studied predominantly at points situated in the sun vertical under cloudless skies.

Already the earliest observations showed that the degree of polarization at any point was highly variable in time and depended both on the state of the sky and on the meteorological conditions and ground (albedo) at the

observation point. Light from clouds is generally unpolarized. The insignificant polarization observed is due to the atmospheric layer enclosed between the observer's eye and the cloud.

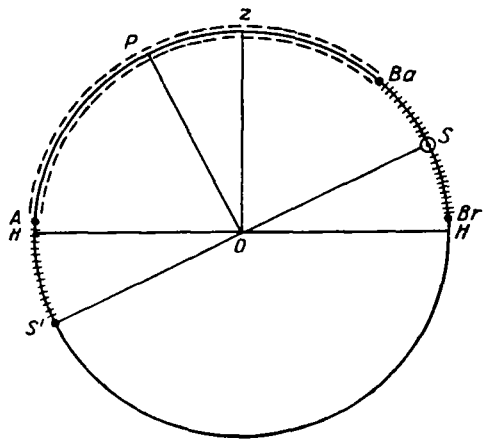


FIGURE 156. Position of neutral points and polarization planes between *A* and *Br*

In the sun vertical (Figure 156) the greatest polarization is always observed at a distance of 90° from the sun (point *P*), where it is usually 60–70% and only exceptionally reaches 85%. The degree of polarization decreases away from this maximum in both directions, and points emitting unpolarized light are recorded at a certain distance (about 18°) above the sun *S* and at the same distance above the antisolar point *S'*; after this polarization increases once again. These are the so-called neutral Babinet point *Ba* and Arago point *A*; a third neutral point, the Brewster point *Br*, lies below the sun at the same distance from it as the Babinet point. The position of the plane of polarization in the part of the sun vertical between the points *A* and *Ba* coincides with the plane of the

TABLE 73

Degree of polarization at points in the sun vertical

$h_{\odot}$(°)	Zenith	Points above the sun (°)				
		50	70	90	110	130
-3	0.75	0.30	0.57	0.75	0.59	0.29
0	0.73	0.17	0.52	0.73	0.54	0.18
5	0.66	0.17	0.47	0.68	0.51	0.20
10	0.57	0.19	0.43	0.65	0.49	0.25
20	0.42	0.18	0.43	0.55	0.44	0.21
40	0.17	0.17	0.36	0.55	0.42	—
60	—	0.18	0.39	0.53	—	—

vertical, i.e., the oscillations of the electric vector are horizontal here Between the point *A* and the antisolar point *S'*, and also between the points *Ba* and *Br*, the plane of polarization is horizontal.

The distance from the neutral point *A* to the antisolar point, as well as the distance from the points *Ba* and *Br* to the sun, do not remain constant; as $h_\odot$ increases the points *Ba* and *Br* draw nearer to the sun while the point *A* moves away from the antisolar point. These distances and the degree of polarization both depend very strongly on the state of the atmosphere.

Data concerning polarization are most complete for the sun vertical. Table 73 lists the mean yearly polarizations in this plane, showing that the greatest polarization is observed at a point situated 90° above the sun. The degree of polarization is usually determined at this point, as well as at the zenith

The degree of polarization is very closely related to the turbidity of the atmosphere, decreasing as the latter increases. Even the slightest increase in the atmosphere turbidity will be accompanied by a reduction of the degree of polarization, making it possible to estimate the transparency of the atmosphere from such observations. Polarization is particularly small for a dust-filled atmosphere and high water vapor content; differences in polarization are therefore observed for different air masses.

TABLE 74

Relationship between degree of polarization and albedo

Albedo	0	0.2	0.4	0.6	0.8	1.0
Degree of polarization	0.852	0.775	0.666	0.612	0.550	0.492

The albedo of the earth's surface also exercises a significant influence on the degree of polarization *p*. Table 74 presents values of *p* as a function of the albedo, from calculations for the point of greatest polarization.

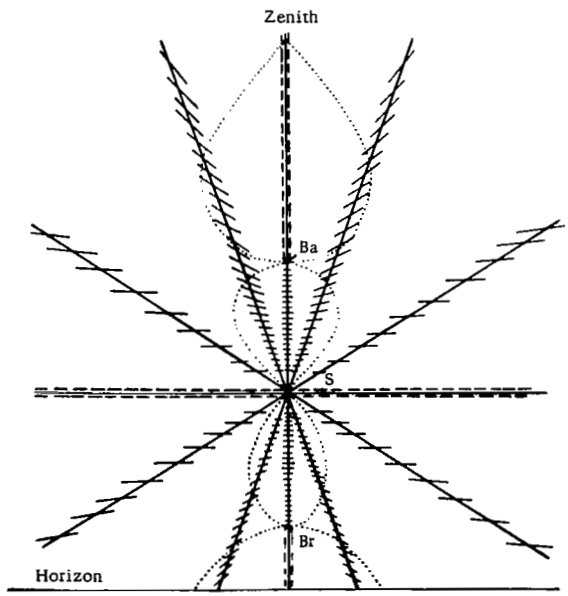


FIGURE 157. Position of plane of polarization around the sun

S–sun; Ba–Babinet point; Br–Brewster point

Polarization is also observed at night when the sky is illuminated by the moon's light.

Other points in the sky lying outside the sun vertical also send off polarized light, the plane of polarization coinciding approximately with the plane of the great circle passing through the sun, the sighted point and the eye of the observer. Hence it lies at a certain angle to the vertical plane (Figure 157). Polarization is conventionally described as positive or negative according as the plane of polarization forms an angle smaller or greater than 45° with the vertical plane

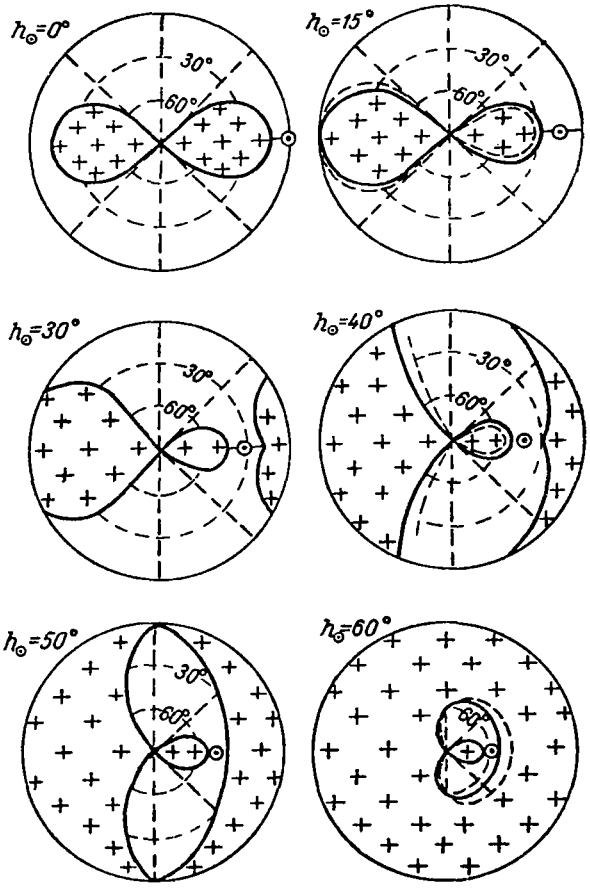


FIGURE 158. Distribution of polarization over the sky for various solar elevations (after A. A. Sinyagin)

By drawing lines through points in the sky at which the plane of polarization lies at the same angle to the vertical plane, we obtain polarization isoclines, as they are known. Isoclines for which the angle of inclination of the plane of polarization is 45° are sometimes called neutral lines. The latter demarcate the boundary between the regions of positive and negative polarization in the sky.

Figure 158 shows the positions of the neutral lines for different $h_{\odot}$, as obtained theoretically by A. A. Sinyagin; the dashed lines represent their positions according to Dorno's observations at Davos (Alps).

§ 6. Daytime illuminance

Direct and scattered rays from the sun illuminate the earth's surface, creating the natural lighting which has such exceptional significance for man and the entire organic world.

When speaking of the illumination of the surface one may consider the illumination produced by direct solar rays or scattered radiation, or else the total illumination. It is usual to measure illumination with reference to a horizontal surface, and the illumination of inclined, variously oriented surfaces is considered only in special cases.

The set of data characterizing the conditions of illumination prevailing at any given point gives an idea of the so-called light climate of this point. The light climate has considerable practical importance, especially for agriculture.

The theoretical data presented for direct and scattered radiation apply in their entirety to the visible regions of their spectra.

The illumination E of a horizontal surface by direct solar rays, which is a function of the zenith distance of the sun, can be written as

$$E = E_0 p_b^{m(z)} \cos z_{\odot}, \tag{8}$$

where p_b is the so-called visual transmission coefficient of the atmosphere and E_0 is the solar illuminance (i. e., the illuminance produced by direct solar rays outside the atmosphere on a surface perpendicular to the sun's rays and placed at the earth's mean distance from the sun). The value of this constant has been less reliably established than that of the solar (energy) constant; it can be taken as 135,000 lux.

The distribution of illumination by direct solar rays over the earth's surface can be calculated theoretically in the same way as that of direct solar radiation; the results will be analogous to those given in Chapter 9

The question of illumination by scattered light can be examined theoretically only for an ideal atmosphere, taking the theory of molecular scattering as point of departure.

Recalling the formula given earlier for the luminance of the sky in a certain direction, one can write down the following expression for the monochromatic illuminance of a horizontal surface due to a certain portion of the sky within the solid angle $d\omega = dz dA$ (where $z_{\odot}$ is the zenith distance of the sun and A the azimuth):

$$E_{\lambda} = b_{\lambda} \cos z_{\odot} \, dz dA.$$

By integrating over the entire sky and over all wavelengths in the visible region of the spectrum one can compute the illumination produced by scattered light from the whole sky in the entire visible spectrum. Such calculations are highly complex and many attempts have been made to simplify this expression and allow for the influence of second-order scattering.

Critical analysis of the numerous formulas puf forward for the illumination under a clear sky indicates that one of the more convenient and accurate ones is that suggested by Makhotkin for scattered irradiance, namely

$$E = \frac{S_0 \cos z_\odot}{1 + \varepsilon_1 \tau \sec z_\odot}, \qquad (9)$$

where ε_1 is the fraction of energy scattered backwards and $\tau = -\ln p$ is the optical thickness of the atmosphere.

TABLE 75

Illuminance at earth's surface (10^3 lux) according to observations at Pavlovsk

Illuminance	$h_\odot$ (°)											
	0	5	10	15	20	25	30	35	40	45	50	55
Direct (10^3 lux)	0.0	1.5	5.8	11.7	17.7	24.6	32.3	39.4	46.2	55.5	63.4	72.4
%	0	36	56	66	71	75	77	79	80	81	82	83
Scattered (10^3 lux)	0.5	2.7	4.6	6.1	7.3	8.4	9.5	10.6	11.6	12.6	13.6	14.5
%	100	64	44	34	29	25	23	21	20	19	18	17
Ratio of direct to scattered	0	0.56	1.26	1.92	2.43	3.00	3.40	3.72	4.00	4.40	4.66	5.03
Total	0.5	4.2	10.4	17.8	25.0	33.0	41.8	50.0	57.8	68.1	77.0	86.9

Turning to the results of observations relating to the illumination of the earth's surface, we begin by citing data on the dependence of the illuminance due to both direct solar rays and scattered radiation on the height of the sun on clear days (Table 75, from observations over many years at Pavlovsk). As one can see, the dependence on solar elevation is much more prominent for illumination by direct solar rays. When $h_\odot$ changes from

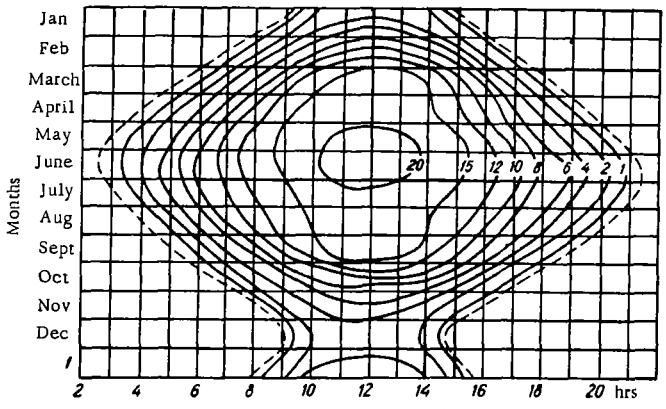


FIGURE 159. Lines of equal illumination (in 10^3 lux) of horizontal surface for natural conditions (Pavlovsk)

Note: Broken lines indicate time of sunrise and sunset.

5 to 55° the latter increases 48.3 times, as against an increase of only 5.4 times in the illumination by scattered light. At low solar elevations the illumination by scattered light is greater than that by direct solar rays.

The daily and yearly periodicities of the illumination are naturally determined for the most part by the dependence on solar elevation. The maximum illumination is recorded in summer in the midday hours, as one can see from Figure 159.

Illumination due to scattered light is especially sensitive to the turbidity of the atmosphere, cloudiness, and the character of the underlying surface. The same regularities are observed in this case as were noted earlier for the entire scattered radiation (cf. Chapter 9).

As follows also from theoretical considerations, with increasing turbidity the illumination by direct solar rays decreases but the luminance of the sky and together with it the illuminance of the earth's surface due to scattered light increase; the total illumination, however, decreases.

The influence of the dust content explains the fact that the total illumination at industrial centers (large cities) is considerably less than in rural localities.

The influence of clouds on illuminance is particularly strong. According to long-term observations at Pavlovsk, the summer midday maximum of illuminance due to scattered light amounts to 12,000 lux for cloudless skies as against 20,000 lux for all days including cloudy ones, i.e., there is an increase of nearly 65%. This increase is a little different for different cloud forms and depends on the cloudiness rating and the solar elevation.

This question, as well as that of the influence of surface albedo on the flux of scattered radiation (illuminance increases rapidly with the albedo), was dealt with in detail in Chapter 9 (§ 3). We might only add that when considering the influence of clouds one must bear in mind that partial cloudiness (provided the sun is not obscured) will increase the total illumination only slightly, while even thin clouds covering the sun will reduce the total illumination very appreciably. Therefore in overcast weather when the sun is obscured by clouds the illumination is usually half of what it would be under clear skies.

Similarly to the question of the illumination of horizontal surfaces, the illumination of inclined and in particular vertical surfaces has very great practical importance. The latter include mountainsides, building walls and other variously oriented surfaces.

The illumination of such surfaces is made up of that due to direct solar rays, the light from a certain part of the sky and the light reflected by surrounding terrestrial objects.

Without pausing over details we note that southern slopes receive a larger amount of direct radiation than western, eastern and especially northern slopes. The ratios between the influxes of radiation to variously oriented slopes vary markedly with the solar elevation. For example, conditions of influx of solar radiation are not always most favorable on southern slopes: this holds, in particular, in the morning and evening hours, especially at high latitudes.

§ 7. Twilight, dawn and dusk. Twilight illumination

For a certain time after sunset and before sunrise the earth's surface receives light scattered by parts of the atmosphere above the horizon still lighted by direct rays.

For any depression of the sun below the horizon (angle α) the height $NM=H'$ at which its rays $SKMB$, tangent to the earth's surface, intersect the plane of the horizon xx, and also the height $AB=H$ at which they pass through the zenith of the observation point A, can easily be computed from geometrical considerations (Figure 160). Indeed, from the triangle OKM one can see that

$$NM=H'=R\left(\sec\frac{\alpha}{2}-1\right). \tag{10}$$

Similarly, from the triangle OKB we obtain

$$AB=H'=R(\sec\alpha-1). \tag{11}$$

We recall that $\alpha=z_{\odot}-90°$ and

$$\cos z_{\odot}=\sin\varphi\sin\delta_{\odot}+\cos\varphi\cos\delta_{\odot}\cos\tau,$$

where φ is the latitude, $\delta_{\odot}$ the sun's declination and τ its hour angle.

It is then easy to calculate the instant of sunrise and sunset at different latitudes, both for the earth's surface and for any height in the atmosphere.

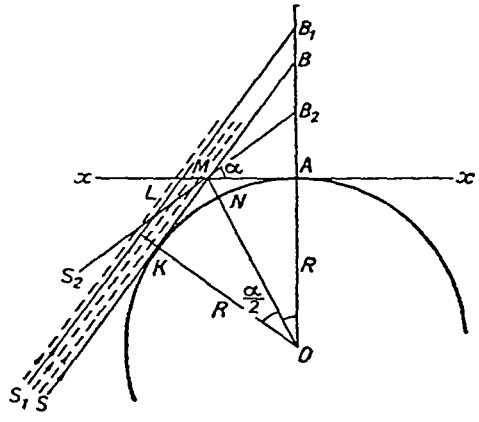


FIGURE 160

Under definite conditions (specified φ and $\delta_{\odot}$) the sun remains above the horizon for a certain time and continuous daylight is experiences ($z_{\odot}<90°$). At the surface this happens in summer at latitudes greater than 66.5°. For higher atmospheric layers the period of illumination by the sun's rays is considerably longer. At heights of more than 500 km above the pole, for instance, the sun does not set throughout the year.

In the evening after sunset as the sun dips gradually below the horizon the illumination of the earth's surface, which is due to scattered light, decreases fairly rapidly until night darkness occurs. The part of the day in which this transition from daytime lighting to nighttime darkness and back takes place is called, correspondingly, evening or morning twilight.

The phenomena observed during morning and evening twilight are the same though occurring in opposite orders; we shall therefore confine ourselves mainly to evening twilight.

While the beginning of evening twilight (and end of morning twilight) is determined exactly by the instant of sunset (sunrise), its termination (beginning of morning twilight) is less definite and is established somewhat arbitrarily. If one observes the fading of twilight in the evening one will notice a point where the illumination is so small as to make any outdoor work without artificial lighting difficult. This instant of practical importance defines the end of the so-called civil twilight. Under cloudless skies it will occur when the sun dips 6–8° below the horizon. As the sun dips further twilight continues until, finally, the brightness of the sky (and therefor the illumination) is so small that astronomical observation of even the faintest stars becomes possible.

This instant is conventionally taken as the end of the so-called astronomical twilight: the sun drops by about 18° below the horizon and the last traces of twilight (cf. below) vanish from the sky.

Obviously, for the above definition the duration of twilight at different points of the earth's surface at different times will be determined by astronomical factors (sun's declination and latitude). Standard tables published in astronomical handbooks or nomograms are used to calculate the twilight duration.

Numerical data for the twilight durations at different latitudes are presented for illustration in Table 76.

TABLE 76

Twilight duration (c- civil, a- astronomical) at various latitudes, in minutes

Latitude (°)	1 January		1 April		1 July		1 November	
	a	c	a	c	a	c	a	c
Equator	76	24	70	23	76	24	70	23
20	80	26	75	24	85	26	74	24
40	99	33	94	30	124	37	92	30
60	168	63	161	54	Continuous	120	146	45

The table shows that the shortest twilights are observed at the equator, where the duration remains practically the same through the year. With increasing latitude the twilights grow longer, especially in summer. In this season the end of the evening and beginning of the morning twilight converge as one moves polewards, and starting from a certain latitude the two may merge for a certain period, resulting in a night-long twilight— the so-called white nights.

For the twilights to merge it is necessary that the following inequality be satisfied

$$90° - \delta_{\odot} - \varphi < \alpha_{\odot},$$

here $\alpha_{\odot}$ is the angle by which the sun dips below the horizon at the instant of the twilight's end.

By applying this inequality one finds that white nights are possible at latitudes greater than 48.5°. They are particularly bright at latitudes greater than 59°, where morning and evening civil twilight merge. At latitudes greater than 66.5° the continuous polar day is experienced in

summer for a certain period of the year while very prolonged twilights and continuous polar night are observed in winter.

In the above distinction between civil and astronomical twilight and their distribution no account was taken of the state of the atmosphere, on which twilight illumination is greatly dependent; we shall proceed to deal with this question.

Table 77 gives mean numerical values for the illumination of a horizontal surface in lux.

TABLE 77

Mean twilight illumination of horizontal surface in lux (after Bullrich)

Height of sun $h_\odot$	Cloudiness		
	0–2	3–5	9–10
5	5000	4700	1500
1	980	750	160
0	550	420	75
–1	280	220	31
–4	15	13	1.4
–6	1.6	1.5	0.17
–7	0.54	0.50	0.064
–8	0.19	0.18	0.026
–10	0.031	0.028	0.006
–16	0.0018	0.0009	0.0005

Cloudiness has a particularly strong influence on twilight illumination. In the early stages of twilight clouds, especially high ones, may sometimes enhance, rather than reduce illumination by reflecting light to the surface. In the middle and final stages clouds usually reduce the illumination, often very considerably so. A continuous mantle of low clouds may reduce illumination by a factor of 10. Snow tends to increase twilight illumination to some extent. To illustrate this we present results of certain observations at Sverdlovsk in Figure 161 (as obtained by V. A. Berezkin). In this figure the curves indicate the illuminance for the following conditions: curve 1 — Ci and As, cloudiness 0–5, no snow cover; curve 2 — same, in the presence of snow cover: curves 3 and 4 — solid St cover, in the absence and presence of a snow cover, respectively.

The following conclusions were arrived at on the basis of observations and calculations:

1) at the beginning of twilight the illuminance amounts to several hundred lux and may reach up to about 500 lux; by the end of civil twilight it drops to roughly 1 lux and, continuing to decrease, by $h_\odot = -10°$ reaches 0.03 lux (i. e., decreases by nearly 10,000 times);

2) the variation of the illuminance at twilight does not depend on the nature of the clouds and underlying surface; the logarithm of the illuminance varies in proportion to the solar elevation (for variation of the latter from –10 to –1°);

3) the illumination may change by a factor of several times according to the cloudiness.

4) the presence of a snow cover increases the twilight illumination.

The elementary interpretation of twilight set forth at the beginning of this section calls for further theoretical elucidation. We shall briefly

present the essence of the twilight theory developed by Academician V. G. Fesenkov and N. M. Shtaude. We note firstly that as the zenith distance of the sun increases the atmospheric layers are extinguished one by one starting from the lowermost; therefore the changes in sky brightness observed during twilight should reflect the peculiarities of atmospheric structure along the vertical. The first person to point out the possibility of utilizing photometric observations relating to twilight for the exploration of the upper atmospheric layers was Fesenkov. Fesenkov also drew attention to the fact that the brightness of the twilight sky at every instant was determined chiefly by light scattering in a definite atmospheric layer occurring at a certain height above the earth's surface.

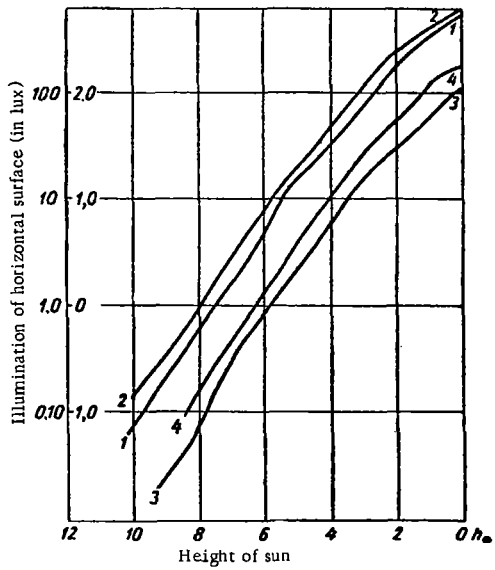


FIGURE 161. Illumination of horizontal surface at twilight, from observations at Sverdlovsk (after V. A. Berezkin)

1-cloudiness 0/0, 3/0, 4/0, 5/0 without snow; 2-same with snow cover; 3-cloudiness 10/10, St, Ns, without snow; 4-same with snow cover.

The appearance of such a layer of maximal contribution to scattered light is easily explained from the following simple considerations. As we saw, the intensity of the light scattered by a certain volume of gas is proportional to the number of scattering particles and to the intensity of the direct solar rays illuminating the given volume. But the flux of the rays depends on the layer of atmosphere traversed by them. Clearly, for the given depression of the sun the flux will increase with increasing elevation above the ground, whereas the number of scattering particles will decrease. As a result at a certain height there ought to be a layer yielding a more or less clear-cut maximum of scattering. This layer will basically determine the observed luminance of the twilit sky.

The appearance of a layer of maximal scattering is observed already some time before sunset, i. e., judging by physical conditions the phenomenon of twilight begins before sunset. For example, for a solar zenith distance of 97° the total twilight luminance is predominantly due to layers lying between 40 and 70 km. The height of these most active layers increases as the sun dips below the horizon, an effect one can regard as due to the rotation (as the zenith distance of the sun increases) of a comparatively narrow beam of solar rays at a roughly constant distance (about 20 km) from the earth. This narrow beam of rays is known as the "twilight ray", and all twilight phenomena can be imagined as the result of the atmosphere's illumination by the twilight ray since, as it happens, rays traversing the atmosphere above or below it have very little effect on the twilight display. Figure 160 shows the position of the twilight ray S_1B_1 and S_2B_2 for two different positions of the sun.

The introduction of the concept of the twilight ray allowed one to obtain formulas for the luminance of the twilight sky in any direction and relating it to the density of the atmosphere at the height of the scattering maximum, the position of which is defined as the intersection of the twilight ray with the line of sight. The theoretical luminance computed from these formulas agrees with the observational results. Subsequently I. A. Khvostikov supplemented luminance measurements by observations of the polarization of the twilight sky, thereby greatly increasing possibilities for exploration of the upper atmosphere.

Photometric and polarimetric observations of the twilight sky make it possible to estimate the pressure and temperature of the atmosphere to heights of about 200–250 km. They constitute a simple and at the same time fairly reliable method of upper-atmosphere investigation ("twilight method").

The method of projector sounding, as it is called, is another optical method of atmospheric research. Essentially, a high-intensity projector beam is aimed upwards and observed from a point a certain distance away from the projector. The intensity, polarization and spectral composition of the scattered light can be studied at different points on the beam. These data also enable one to draw important and varied conclusions concerning the physical properties and state of the atmospheric layers that scatter light. With a sufficiently intense beam of light one can sound the atmosphere from the ground to heights of the order of several tens of kilometers.

In addition to the luminance changes just discussed, across a large portion of the twilight sky (mainly where the sun lies hidden below the horizon) one may also observe a characteristic succession of colors as well as other color phenomena displaying a wide variety of intensities and forms. These optical phenomena are called d a w n - a n d - s u n s e t p h e n o m e n a, or simply dawn and sunset. Despite their great diversity one can single out a number of outstanding characteristic features.

When the sun draws close to the horizon before setting it gradually takes on an orange and often red hue. At this time the sky near the sun and also on the opposite side from it begins to color. Clouds present in the sky likewise become tinted, low clouds assuming a red hue, middle clouds an orange or yellow hue and high clouds remaining white as before. Clouds occurring in the region opposite to the sun color first. As the sun approaches the horizon the intensity of the coloring increases and when it vanishes from sight a bright orange or red patch remains. Color bands

begin to form on both sides of the setting sun, faintly at first but deepening in hue with time. The set of these color bands constitutes the sunset (dawn) segment and is termed the sunset (dawn) proper. In these bands the colors, moving away from the horizon, succeed each other in the order of the colors of the spectrum: red, orange, yellow, green, blue. Some may be missing occasionally but the order is always the same. The uppermost part of the sunset (dawn) is either whitish or bluish. Immediately after the sun has set a dark ash-colored segment, easily distinguished from the rest of the sky, appears in the opposite side; it represents the earth's shadow, and spreads gradually upwards. Eventually the sunset segment shrinks, the clouds fade, all sky colors dim and by the termination of civil twilight the sunset has ended. After this one may sometimes observe a short-lived and far less intensive second sunset glow, followed by a whitish color display consisting of a small yellowish segment; but this too ulitmately vanishes—astronomical twilight comes to an end. Such, schematically, is the development of the sunsets and sunrises observed in the European USSR.

In mountain regions, especially in the presence of snow-covered peaks, the mountains often take on a special coloring at twilight. In fair weather snowy crests usually acquire a reddish tint before sunset which vanishes at the instant of setting: soon after sunset, however, the peaks and slopes again take on a bright red color, usually starting some few hundred meters below the crestline and spreading gradually upwards. Presently the color effect fades once more but may after a brief pause reappear, this time with a yellowish or purplish hue and much less intense. This phenomenon is known in the literature as "Alpine glow". In the USSR it is frequently seen in the Caucasus.

The connection between the character of the sunset and the weather was discovered long ago. However, the extreme diversity of the associated phenomena has so far hampered an exact solution of this problem.

Observations definitely show that the intensity of all sunset (dawn) phenomena increases with the concentration of dust and water vapor in the atmosphere. For instance, especially intensive and colorful displays are observed when the atmosphere is contaminated by smoke from forest and turf fires, or after violent volcanic eruptions. Therefore it can be assumed that the observed diversity of sunset (dawn) phenomena and their relationship to the weather are chiefly determined by differences in the atmosphere near the observation point.

No complete theory of sunset (dawn) phenomena embracing both the quantitative and the qualitative aspects is yet in existence, a situation which can be ascribed to the very great number of factors that must be taken into account. Qualitatively the origin of many of the phenomenon can be understood in the light of our previous discussion of twilight, taking into consideration the attenuation and scattering of light in the atmosphere and especially the influence of atmospheric turbidity.

Together with photometric observation of the twilight sky illumination and polarization, the study of the twilight spectrum is very valuable. Generally speaking the twilight spectrum resembles the solar spectrum with somewhat strengthened telluric lines and absorption bands. But under definite conditions it often displays two special lines: a red line ($\lambda = 6300$ Å) belonging to atomic oxygen and a bright yellow sodium ***D*** line ($\lambda = 5893$ Å). The question of the spectrum of proper emission of the atmosphere will be considered in greater detail later.

§ 8. Nighttime illumination. Illumination due to moonlight. Nightglow

At the end of astronomical twilight on clear, moonless nights the earth's surface receives light from the sky, the luminance of which, evenly distributed over its surface, amounts to about $1.8 \cdot 10^{-3}$ apostilb. This number is smaller for a sky partially or completely covered by clouds. Such an evenly glowing sky will produce an illumination of about $1.8 \cdot 10^{-3}$ lux on a horizontal surface. This extremely weak nightglow is suppressed on moonlit nights by the light sent off by the moon.

The illumination produced by moonlight varies over a wide range, primarily in response to the fact that the intensity of moonlight varies in accordance with the periodic changes in the form of the visible portion of the lunar disk (phases of the moon). Furthermore, the illumination changes as the earth's distance from the moon and moon's distance from the sun change. All this makes calculation of lunar illumination fairly complicated. The fundamental quantity is the so-called light constant for the full moon. This is the illumination produced outside the atmosphere by the full moon at the mean distances of the moon from the earth and sun; its numerical value can be taken as 0.291 lux. Lunar illumination varies with the phases of the moon, not in proportion to the luminous surface of the disk but rather much more rapidly and somewhat differently for the waxing and waning moon, so that in the first and last quarters, for instance, it amounts to only about 0.04–0.03 lux. If data for the intensity of moonlight outside the atmosphere for different phases are available, the lunar illumination can be calculated; to do this, in addition to considering the astronomical factors noted above, one must also allow for the influence of the terrestrial atmosphere. In the first approximation this can be done in the same way as for the daytime illumination, allowing for the lower intensity of moonlight. To obtain more exact data it must be considered that the spectral composition of moonlight is somewhat different from that of sunlight. Such calculations lead to the same regularities as for daytime illumination. They show that for average conditions of atmospheric transmission the full moon at the zenith gives an illumination close to 0.25 lux on a plane normal to its rays. More detailed data for the full moon are given in millilux in Table 78.

TABLE 78

Illumination of horizontal surface by moonlight at full moon, for cloudless weather

Illumination (10^{-3} lux)	Lunar elevation (°)						
	10	20	30	40	50	60	70
Total	21	62	110	158	201	239	267
Scattered light	13	17	22	26	29	33	37

As the data in the table show, the illumination due to moonlight appreciably exceeds the nighttime illumination and also the illumination at the end of twilight when the sun is more than 10–12° below the horizon. During civil twilight the lunar illumination does not play an important role in the general lighting of the earth's surface and may be disregarded.

On moonless nights the surface receives the feeble sky light mentioned in the beginning of this section.

Several sources of light of the night sky can be noted, including: 1) direct and scattered starlight, 2) zodiacal light, chiefly from the zone of the zodiac, 3) galactic light, from nebulae and the Milky Way, 4) self-luminescence of the night sky (nightglow), due to the glowing of gases in the upper atmospheric layers, and, finally, 5) the scattered light from the last three sources.

Despite the difficulty of breaking down the light of the night sky into its individual components, one can roughly assume that the atmospheric component accounts for about 25%.

From the geophysical point of view the nightglow presents the greatest interest. Its photometric and spectroscopic study make it possible to explore the upper atmosphere to very great altitudes.

The nightglow spectrum consists of a system of numerous lines and bands superposed on the continuous spectrum. The intensity of the continuous spectrum is especially marked in the blue and ultraviolet regions, which contain a series of absorption lines coinciding with the Fraunhofer lines in the solar spectrum.

Among the lines and bands of the emission spectrum the most characteristic and intense are: 1) green ($\lambda = 5577$ Å) and red ($\lambda = 6300$ Å and $\lambda = 6363$ Å) lines belonging to atomic oxygen; 2) intense yellow sodium lines ($\lambda = 5890$, 5896 Å); 3) a series of bands of molecular nigrogen, including the Vegard-Kaplan bands (λ about 4420 and 4200 Å); 4) intensive infrared emission in the wavelengths 8000–11,000 Å (especially bright near $\lambda = 10,440$ Å), which correspond to vibration-rotation bands of the hydroxyl molecule; the intensity of these bands is fairly high and exceeds that of the green line of oxygen by about 100 times.

In addition to the above-mentioned lines and bands there is a series of fainter bands belonging to the nitrogen ion N^+, the nitrogen molecule N_2 (first positive system, extending from the green to the infrared region), and the oxygen molecule O_2 and other molecules (H_2O, NO, O_3, etc.).

The spectrum of nightglow is similar in many respects to the auroral spectrum, as we shall see, but there are certain differences confined mainly to the degree of excitation, which is lower for nightglow.

When the spectrum of light from the night sky is compared with the spectrum of scattered light from the daytime and twilight skies, one finds that twilight is marked by a considerable intensification (by a factor of 10–100) of such bands as the yellow sodium lines, red lines of O_2 and hydroxyl lines; this strong enhancement of the intensity justifies the apellation "twilight flash" given to this phenomenon.

Without going into the highly complex mechanism of luminescence of the night sky, certain aspects of which have yet to be cleared up, we note that the energy of luminescence is released in the recombination of atoms formed in daytime as a result of the dissociation of molecules under the influence of ultraviolet emission from the sun.

The height of the atmospheric layers in which this emission takes place has been determined repeatedly by many investigators. It is somewhat different for different lines and moreover the values obtained display considerable scatter: the only thing that is certain is that sodium lines are emitted from heights of 35 to 65 km (twilight flash) and also from heights of about 250 km (from night sky spectrum). As to the oxygen lines and

especially the green one, the height of the layers that emit them is highly variable right up to the upper limits of the atmosphere.

In addition to spectroscopic studies of nightglow a considerable volume of material from photometric measurements of the intensity, both in total and in individual emission lines is available. The data show that the luminance of the glow is different in different sectors of the sky and increases generally towards the horizon, with isolated regions of particularly high luminance. The luminance also appears somewhat different at different points. At the same time the luminance of the night sky and hence the nighttime illumination do not remain constant but change from night to night as well as during the night; this is why certain nights are particularly light.

Variations in time also exhibit diversity. Thus annual fluctuations are prominent. To illustrate the annual march we present in Figure 162 the intensity variations in: a) the green line of O (λ=5577 Å), b) the blue region of the spectrum (Vegard-Kaplan band) and c) the red region, according to observations in England (Terling, 43° from north magnetic pole), Australia (Canberra, 38° from south magnetic pole) and South Africa (Capetown, 69° from south magnetic pole).

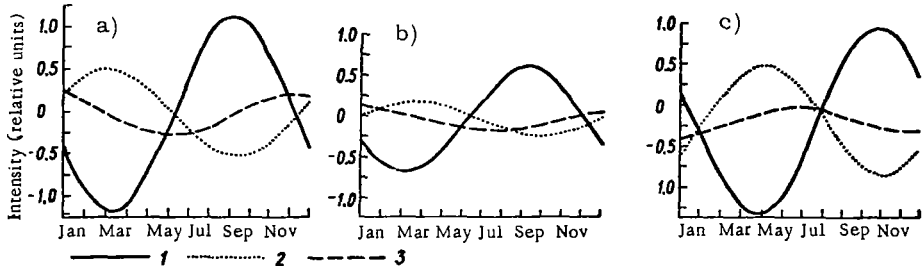


FIGURE 162. Annual march of intensity in green (a), blue (b) and red (c) lines of the nightglow spectrum

1-Terling; 2-Canberra; 3-Capetown.

The figure shows that at each station the intensity variations in the three regions of the spectrum are similar in phase; the largest amplitude occurs in the green region and the smallest in the blue region; furthermore, the smallest amplitude is observed at the station farthest from the magnetic pole (Capetown).

The daily periodicity is fairly difficult to establish owing to the presence of frequent irregular changes (fluctuations), but on the basis of observations on "quiet" nights when fluctuations are small one can single out two basic types of daily variation: one with a brightness minimum at about midnight and the other displaying the opposite character, with maximum about midnight; this is apparently due to differences in the height of the atmospheric layers responsible for emission.

Long-term observations indicate the presence of cyclic variations in the luminance of the sky roughly parallel to the eleven-year cycle of solar activity; these are especially sharp for the green line. Furthermore, there is a tendency towards a 27-day frequency in the intensity variations of the green line.

Finally, the intensity of nightglow displays irregular changes which in the case of the emission of atomic oxygen have been found to be closely related to magnetic activity. This also accounts for the fact that the mean intensity of glow is a little higher at high latitudes. Correlations between irregular flashes of airglow and chromospheric eruptions on the sun, and between the former and irregular processes in the ionosphere (cf. Chapter 29) have also been established definitively.

§ 9. Luminance, color and reflectance of objects of the natural landscape

The question of the luminance and color of landscape objects is closely related to the illumination of the earth's surface. The flux of radiant energy from natural light sources (sun, moon, sky) reaching the surface is partly absorbed and partly reflected by it. Different objects on the surface and different parts of the landscape display different luminances and, in the visible region of the spectrum, different colors depending on the properties of reflecting surfaces. Two types of reflection are distinguished according to structural features of the surfaces: mirror reflection, taking place according to Descartes' laws (equality of angles of incidence and reflection), and diffuse reflection by matt surfaces, in which light is scattered in all directions. In nature one is usually dealing with mixed reflection, in which part of the reflected stream is directed as in mirror reflection, and part as in diffuse reflection. The simplest way of describing the reflecting properties of a surface is by specifying its reflectance or albedo (Chapter 8).

When studying the luminance of objects it is convenient to replace the reflectance by the luminance factor r, which is the ratio of the luminance B of a given surface to the luminance B_w of a perfectly matt white surface observed in the same direction and under the same lighting conditions $r = B:B_w$, or to the illuminance E of the surface, i.e.,

$$r = \frac{\pi B}{E} \text{ stilb/phot} = \frac{B}{E} \text{ lambert/phot.}$$

The luminance factor r is constant only for a perfectly matt (orthotropic) surface for which the luminance is the same in all directions and is proportional to the illuminance. For actual surfaces which are not perfectly matt (nonorthotropic) the luminance factor depends on the angle of incidence i, the angle of reflection j and the azimuth of the reflected ray $A: r = f(i, j, A)$.

This quantity varies widely for different objects and is fairly complex. It can be expressed in the form of a vector diagram, marking off line segments in various directions from the origin so that the lengths be proportional to the values of r in these directions. The geometric locus of such vectors gives the surface called the reflection indicatrix.

In view of this circumstance in most cases the luminance factor is determined in the direction forming an angle of 45° to the surface under investigation and at an azimuth of 90° with respect to the sun, so as to allow comparison of observational results.

For a more complete characterization of the reflectance of natural formations one must recall that they reflect differently in different wavelengths

and hence acquire various hues. Therefore it is necessary to study the so-called spectral luminance factors $r(\lambda) = b(\lambda) : e(\lambda)$, performing measurements in narrow spectral intervals $d\lambda$. If the spectral composition of the illumination is known and one has the values of $r(\lambda)$, one can determine the color of an object.

The most advanced method of determining the spectral luminance factor is the spectrophotometric method, but in addition measurements are often carried out with the aid of photosensitive elements and thermoelectric pyranometers equipped with filters for separating the individual spectral regions.

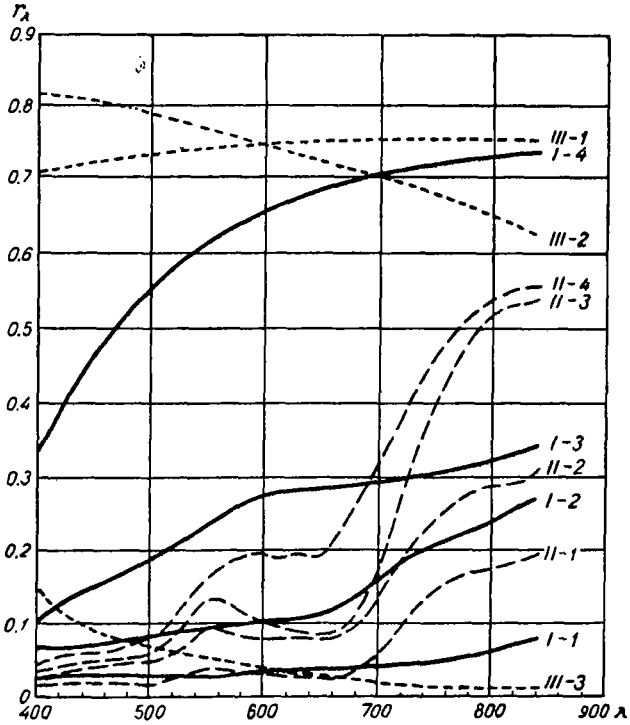


FIGURE 163. Spectrophotometric classification of natural formations (after E.L. Krinov)

Particularly extensive researches into the spectral reflectance of natural formations were conducted by E. L. Krinov. The large volume of factual material he obtained enabled Krinov to divide the reflection curves into three main classes despite the diversity of forms involved, and further to subdivide each of these into several (up to four) types according to the degree of prominence of the criteria characteristic of the given class. This classification is shown in Figure 163. The first class comprises curves in which the luminance factor increases continuously and more or less smoothly from the violet to the red end of the spectrum (solid curves). Reflection of this kind is typical of bare soils and also of structures of various kinds. The second class is characterized by a maximum of reflection

in the visible region near $\lambda = 560\,m\mu$ and a very high reflectance in the far red and infrared regions (dashed curves); such reflection is noted for plant covers (forest, field, and so on) in the growth period. Finally, the third class includes curves either exhibiting no marked change in the reflectance with changing wavelength, or in which the luminance factor decreases from the violet to the red end of the spectrum (dotted curves). Such reflection is typical of snow covers and water surfaces.

We note further that moist surfaces have a lower luminance than dry ones, the constrast in luminance being sometimes very considerable (factor of three).

Chapter 25

VISIBILITY

If any distant object (structure, landscape element, etc.) is observed systematically it may be noticed that at certain times the object is very clearly visible while at other moments, on the contrary, it becomes so blurred that only its contours and large components are distinguishable; occasionally it may even become completely invisible. In such cases one speaks of good or poor visibility. The question of visibility has great practical significance, in particular for all forms of transportation and especially for aviation. For meteorology visibility observations are valuable as a method for studying atmospheric transmission, since the optical properties of the atmosphere are one of the principal factors governing visibility conditions.

§ 1. Visual range. Principal factors governing visual range

Visibility is characterized numerically by the so-called visual range, which can be defined:

a) either as the greatest distance at which the eye can still detect the first signs of the presence of a given object,

b) or as the smallest distance starting from which the last signs of the presence of a given object disappear and it becomes impossible to determine its position against its background.

The former is called the detection distance of the object and the latter the distance of visibility loss of the object. The former, obviously, is always smaller than the latter. The difference between them determines a certain interval of distance in which the visibility of the object becomes unreliable— this is the zone of uncertain visibility.

Visual range is usually measured in kilometers and, for very poor visibility as in fog, in meters.

It has been established that the visual range depends on a large number of factors. The principal ones are:

1) optical properties of the atmosphere. These determine, on the one hand, the attenuation of the luminous flux from the object and background to the eye of the observer, and, on the other hand, the intensity of the scattered light which reaches the observer's eye from air layers occurring between the object and the observer and which produces the so-called air light;*

* [Lit. "air haze" in the Russian.]

2) properties of the object sighted, its angular dimensions, form, color and in particular its photometric characteristics (reflectance, etc.);

3) properties of background against which the object is examined— its luminance, color, reflectance, etc.;

4) conditions of lighting of object and background;

5) properties of apparatus with which observation is performed— the optical characteristics of the instrument and especially the properties of the observer's vision (sensitivity of the eye to the reception of luminance, color, contrast, its resolving power, etc.).

§ 2. Effect of the atmosphere on the apparent luminance of an object. The air light equation

Let us elucidate the effect of the atmosphere on the observed luminance of a sighted object. Consider a screen P subtending the solid angle $d\omega$ at the sighting point O, which is at a distance l (Figure 164). If the intrinsic

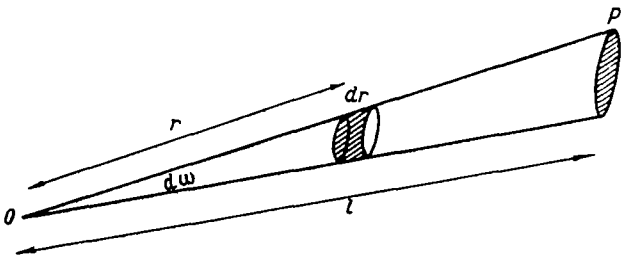


FIGURE 164

luminance of the screen is B_0, then the luminous flux from it will reach the eye weakened in the ratio $e^{-\alpha l} = p^l$, where α is the light attenuation coefficient and p is the transmission coefficient. Therefore the luminance of the screen as seen by the eye will be not B_0 but

$$B_1 = B_0 e^{-\alpha l} = B_0 p^l. \tag{1}$$

At the same time the air layer situated between the eye and the screen within the cone OP is illuminated both by direct solar rays and by light scattered from the sky and the earth's surface; as a result this layer also has a certain luminance. To simplify the calculations we will assume that the optical properties and conditions of illumination are identical at all points of the layer. Let us take the volume element $dv = r^2 d\omega dr$ at the distance r from the eye O (Figure 164). This element will scatter light in the direction of the eye. Its luminance $d\beta$ in the direction of the observer is the luminous intensity per unit solid angle due to it at the point O, i.e.,

$$d\beta = \frac{cE dv}{r^2 d\omega} = cE dr, \tag{2}$$

where c is the scattering coefficient and E is the illuminance of the volume element.

But as perceived by the eye this luminance will be attenuated in the ratio $e^{-\alpha r}$, and the apparent luminance of the volume element will be

$$(d\beta)_1 = d\beta e^{-\alpha r} = cEe^{-\alpha r}\,dr. \tag{3}$$

The luminance of the entire air layer contained within the cone between the eye O and screen P can be obtained by integrating the above expression over r from 0 to l. This gives

$$\beta_1 = \int_0^l cEe^{-\alpha r}\,dr = \frac{cE}{\alpha}\left[1 - e^{-\alpha l}\right] = \frac{cE}{\alpha}\left[1 - p^l\right]. \tag{4}$$

Introducing the notation $\frac{cE}{\alpha} = B_h$, we obtain

$$\beta_1 = B_h\left(1 - p^l\right) = B_h\left(1 - e^{-\alpha l}\right). \tag{5}$$

The above gives the luminance perceived by the eye from the air layer contained between the eye and the object under consideration— the so-called air light luminance. The above shows that the greater the distance to the object the greater this luminance; it also increases with increasing turbidity of the atmosphere (increasing α).

Thus as a result of the influence of the atmosphere the apparent luminance of the object is not B_0 but

$$B_0 = B_1 + \beta_1 = B_0 e^{-\alpha l} + B_h\left(1 - e^{-\alpha l}\right) = B_0 p^l + B_h\left(1 - p^l\right). \tag{6}$$

This is known as the air light equation. The quantity B_h appearing in it is called the air light coefficient. The latter can be interpreted as the luminance of an infinite air layer in the direction to the object for specified observation conditions. Indeed, setting $l = \infty$ in equation (5) we obtain $\beta_{1,\infty} = B_h$. In practice this limiting value of the air light luminance depends on the lighting conditions and on the turbidity of the atmosphere; it is often taken to be equal to the luminance of a cloudless sky at the horizon. Frequently when the atmospheric turbidity is high (large α) the limiting value of B_h is obtained already at comparatively small values of l. In fog, for example, l amounts to a few tens of meters in all and represents the luminance of a "wall of fog" obscuring the horizon.

Expression (6) shows that the apparent luminance of an object, B_0' approaches the quantity B_h as the distance to it increases, and therefore if $B_0 < B_h$ (i.e., the object is darker than the horizon sky) the apparent luminance of the object will increase as one moves away from it— the object will become brighter; for $B_0 > B_h$ (the object is brighter than the horizon sky), on the contrary, it decreases as one moves away— the object becomes darker

§ 3. Luminance contrasts. Visual acuity

A sufficiently large object can be detected only if it differs in color or luminance from the background against which it is perceived. This difference is characterized by the so-called color or luminance contrast, respectively. It should be noted that under real conditions one is dealing with mixed contrast, which is a combination of the differences in color and

luminance between the object and its background. Since in practice, however, the luminance contrast is decisive in most cases, we shall henceforth confine our attention to it. The luminance contrast K is defined as the ratio of the absolute difference in the luminances of the object B_o and background B_b to the greater of the two, i.e.,

$$K = \frac{B_o - B_b}{B_o}, \quad \text{if} \quad B_o > B_b,$$

or

$$K = \frac{B_b - B_o}{B_\phi}, \quad \text{if} \quad B_b > B_o. \tag{7}$$

Obviously, the contrast does not depend on the absolute values of B_o and B_b and, by definition, is always a positive proper fraction, i.e., $0 \leqslant K < 1$; its limiting values are $K=0$, if $B_o = B_b$, and $K=1$ if the lower luminance is zero, i.e., when either the object or the background can be treated as a black body. In the first case ($K=0$) if the background and object are the same color the object will not be recognizable, whatever be its luminance.

In distinguishing contrasts with our eyes we cease to perceive them a little before K becomes zero. The smallest contrast for which the eye ceases to perceive it is called the threshold of brightness contrast of the eye and is denoted by ε. Having introduced it one can say that an object will be visible only when $K > \varepsilon$ and not visible when $K < \varepsilon$; the equality $K = \varepsilon$ determines the state of transition from visibility of the object to its invisibility.

In addition to being governed by the contrast threshold of the eye, the possibility of perception of an object against a background is also determined by the acuity of vision, which is the ability of the eye to distinguish two points placed one next to the other. Numerically visual acuity is characterized by the inverse of the so-called smallest angle of resolution δ, expressed in minutes; the latter is the limiting angular distance between two points at which they can still be distinguished as separate, i.e.,

$$v = \frac{1}{\delta}. \tag{8}$$

The standard of visual acuity is conventionally taken as its value for δ equal to one minute of arc. The arbitrariness of this quantity is very considerable, since visual acuity while being a characteristic of the observer also depends on many factors such as the form of the object, the illuminance level, contrasts between object and background, and so on. If when observing an object of angular dimensions γ the limiting contrast for which it can still be distinguished is K, then it so happens that the following approximate relationship holds between the contrast K and the angle δ:

$$(K - \varepsilon)\gamma^2 = (1 - \varepsilon)\delta^2, \tag{9}$$

this shows that if $K=1$ the eye will be able to resolve an angle equal to the smallest angle of resolution $\delta = \frac{1}{v}$. If the angular dimensions of the observed object are greater (angle γ is large) the above gives us

$$K - \varepsilon = \frac{(1 - \varepsilon)\delta^2}{\gamma^2}. \tag{10}$$

From this it follows that with increasing γ the difference $K - \varepsilon$ decreases, i.e., the least resolved contrast K approaches the threshold of brightness

contrast ε. This leads to the necessity of using objects with sufficiently large angular dimensions when determining contrast.

The threshold of brightness contrast ε is not invariant. In addition to being somewhat different for different observers, its value depends strongly on conditions of observation and most of all on the illuminance and angular dimensions of the object. With decreasing illuminance ε increases rapidly while visual acuity decreases. Therefore at twilight and at night when the illuminance is small the threshold of brightness contrast increases strongly and may reach up to several tens of percents (ε=0.6–0.7); this means that objects distinguishable from the background in daytime become indistinguishable at twilight even though the contrast betweem them and their background remains the same.

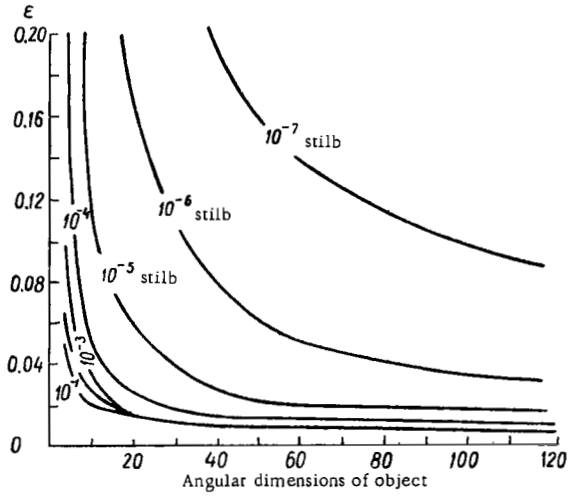


FIGURE 165. Threshold of brightness contrast of the eye ε as function of angular dimensions of object for different background luminances

Figure 165 illustrates the dependence of the threshold of brightness contrast of the eye ε on the angular dimensions of the object (in minutes of arc) for different background luminances. As one can see, under daytime conditions ε remains practically constant for objects with angular dimensions exceeding 20' and does not depend on the illuminance and dimensions of the object; under nighttime conditions and at twilight the dependence becomes very clear-cut.

The magnitude and constancy of the threshold of brightness contrast and visual acuity are affected further by a whole series of factors such as the degree of adaptation of the eye to the prevailing conditions of illumination, the presence of subsidiary light stimuli and the psychological state of the observer. This should always be borne in mind when performing observations of the visibility.

The following values of ε and v (and correspondingly δ) have been adopted on the basis of numerous special studies for practical calculations of the visual range for visual observations in the daylight hours:

1) for determining the distance of visibility loss

$$\varepsilon = 0.02, \qquad v = 2.0 \qquad (\delta = 0.5');$$

2) for determining the detection distance

$$\varepsilon = 0.05, \qquad v = 1.5 \qquad (\delta = 0.7').$$

§ 4. Fundamental theorems of the theory of visual range

The principal aim of the theory of visibility is to find the functional dependence between the visual range and all the factors that determine it. This task is very complicated and we shall confine ourselves here to considering the basic propositions of the theory with reference to the determination of visual range in a horizontal direction in daytime (daytime horizontal visibility).

Let an object of luminance B_o be observed against a background of luminance B_b. We will assume that $B_b > B_o$. Denoting the distance to the object by l and that to the background by $l+\Delta l$, according to (6) one can express the apparent luminance of the object as

$$B_o' = B_o p^l + B_h(1 - p^l), \tag{11}$$

and that of the background as

$$B_b' = B_b p^{l+\Delta l} + B_h(1 - p^{l+\Delta l}). \tag{11'}$$

The contrast produced by these luminances (if $B_b > B_o$) will be

$$K = \frac{B_b' - B_o'}{B_b'} = \frac{B_b p^{l+\Delta l} - B_o p^l - B_h(1-p^l)}{B_b p^{l+\Delta l} + B_h(1-p^{l+\Delta l})}, \tag{12}$$

or

$$K = \frac{(B_b - B_h) - \dfrac{B_o - B_h}{p^{\Delta l}}}{(B_b - B_h) + \dfrac{B_h}{p^{\Delta l}}\dfrac{1}{p^l}}. \tag{12'}$$

On the basis of the earlier arguments, the visual range is given by that limiting value $l = L$ at which the contrast K equals the threshold or brightness contrast of the eye. This value L can be obtained from (12) by setting $K = \varepsilon$

$$L = \frac{1}{\ln p} \ln \frac{\varepsilon}{\left(1 - \dfrac{B_o}{B_h}\right) - \left(1 - \dfrac{B_b}{B_h}\right)(1-\varepsilon)\, p^{\Delta l}}. \tag{13}$$

The above expression gives the solution of the problem for the case under consideration under the above-mentioned assumptions: identity of the colors of the object and its background, uniformity of the illumination and constant turbidity of the atmospheric layer over the distance L. In this expression the ratios $\frac{B_o}{B_h} = b_o$ and $\frac{B_b}{B_h} = b_b$ are the so-called brightness factors of the object and background, respectively.

The brightness factor is an abstratc quantity. For a dark object it is less than unity, for a light object greater than unity; for a black object it is zero. The brighness factor of a cloudless sky at the horizon is unity $(B_b = B_h)$. For a dark cloud the brightness factor will be less than unity; for

a light cloud it will be slightly greater than unity. For a sky uniformly covered by clouds the brighness factor will be close to unity.

Formulas (12) and (13), however, are inconvenient for practical use and are not usually employed in the form given above. They can be simplified by setting $\Delta l=0$, i.e., assuming that the distances to the object and background are the same. Then, introducing the brightness factors b_0 and b_b, after simple transformations we obtain in place of (12)

$$K=\frac{B'_b-B'_0}{B'_b}=\frac{(B_b-B_0)\,p^l}{B_b p^l+B_h(1-p^l)}=\frac{K_0}{1+\frac{1}{b_b}(p^l-1)}, \qquad (14)$$

where $K_0=\frac{B_b-B_0}{B_b}$ is the true contrast between the object and the background and b_b is the brightness factor of the background (if $B_b>B_0$).

If $B_0>B_b$ then, analogously,

$$K=\frac{K_0}{1+\frac{1}{b_0}(p^l-1)}. \qquad (14')$$

If the background is a cloudless sky at the horizon, for which the brightness factor is unity, expression (14) becomes even simpler and assumes the form

$$K=K_0p^l=K_0e^{-\alpha l}. \qquad (15)$$

This equation can also be obtained directly from the air light equation (6). Taking the logarithm of (15) we find that

$$\ln K=\ln K_0-\alpha l. \qquad (16)$$

To determine the visual range L one must set $K=\varepsilon$; then from the above we obtain

$$L=\frac{\ln K_0-\ln \varepsilon}{\alpha}. \qquad (17)$$

If, furthermore, the observed object is a black body of intrinsic luminance $B_0=0$, then (taking $K_0=1$) we find

$$L=S_m=-\frac{\ln \varepsilon}{\alpha}. \qquad (18)$$

The visual range S_m so defined is called the meteorological visual range.

Consequently, the meteorological visual range is the greatest distance at which, for the given atmospheric transmission, a black object of large angular dimensions (over 20') projected against a horizon sky will fuse with its background and become invisible.

From (18) it is seen that if S_m is determined from observations one can find the attenuation coefficient α by taking an appropriate value for the threshold of brightness contrast of the eye ε.

Since by definition the meteorological visual range is the distance of visibility loss for which $\varepsilon=0.02$, we have

$$S_m=-\frac{\ln 0\,02}{\alpha}=\frac{3.91}{\alpha}. \qquad (19)$$

It should be emphasized that S_m characterizes merely the state of transmission of the atmosphere.

The above discussion applies to the daytime horizontal visibility of objects. At twilight, when the illumination decreases strongly, and all the more so at night when it becomes negligible, the solution of the problem of the visual range of landscape objects becomes extremely complicated.

These complications are due only to a change in the visual properties. Indeed, we saw earlier that as illumination fades the threshold of brightness contrast increases sharply while visual acuity decreases. Although the magnitude of the contrast does not depend on the luminances of the object and background, as a result of these changes in the threshold and visual acuity visibility of objects at night is lost at much smaller distances than in daytime. Under twilight conditions when the illumination and with it the threshold of brightness contrast are changing continuously and very rapidly, it becomes impossible to indicate any constant value for the latter and hence to calculate S_m. Owing to this all formulas and calculations based on the assumption that the threshold of brightness contrast is constant are valid for daytime but meaningless for twilight and night.

The meteorological visual range, which is uniquely related to the atmospheric transmission (attenuation coefficient for light α), is one of its characteristics. Determination of the meteorological range is very important for calculations of the visual ranges of real objects. We recall that as one of the principal meteorological elements it is evaluated at meteorological stations according to the ten-point international scale given in Table 79.

TABLE 79

Scale of meteorological visual range

Degree	Object at this distance is	
	visible	not visible
0	—	50 m
1	50 m	200 m
2	200 m	500 m
3	500 m	1 km
4	1 km	2 km
5	2 km	4 km
6	4 km	10 km
7	10 km	20 km
8	20 km	50 km
9	50 km or more	—

Objects placed at the distances indicated in the second column of the table, which satisfy definite requirements corresponding to the conditions of the theory and specified in the Directions, are chosen for performing the observations. However, the applicability of this method is severely limited by the fact that at many stations objects occurring in the observer's field of vision and corresponding to the prescribed conditions (dark, large, projected against horizon sky, etc.) are difficult to find in the required numbers. The relative error of such determinations of S_m is very high. Other auxiliary methods have therefore been developed and recommended for visual estimation of the visual range, in particular a method based on evaluation of the obscuration by air light of objects occurring even at

comparatively small distances from the observer. For nighttime observations single lights of known luminous intensity are placed at definite distances

The difficulties involved in organizing and performing visual observations, which call for long experience on the part of the observer, induced researchers at any early date to devise special instruments for measuring the meteorological visual range. A number of instruments and methods of measurement have now been developed for meteorological stations. In most visibility meters the contrast between object and background observed in the instrument is artificially lowered by some optical device to the threshold of brightness contrast of the eye; details on the use of these instruments and methods of measurement will be found in the corresponding monographs.

§ 5. Visual range of real objects

We noted above that when determining the meteorological visual range the observations must be conducted in strict conformity with the conditions underlying formula (18). Since fulfillment of these conditions is often tedious and complicated one is obliged to depart from them in some respects, especially by substituting dark objects encountered in the real situation for black ones prescribed by the theory. At the same time, in practice it is often required to indicate the visual ranges of specific objects for different states of the atmosphere and at different times of day (different illumination). This problem, however, is not so much meteorological as one of illumination engineering: it is also extremely complex, requiring for its solution above all a knowledge of the atmospheric transmission (meteorological visual range). This raises the important question of the effect of the properties of objects, conditions of illumination and properties of the eye on the visual range as determined from them. Certain of these effects have already been considered. Here we shall further discuss the influence on visual range of the "true" (proper) luminance of the objects, undistorted by air light, which depends on the albedo of the objects and on lighting conditions.

Theoretical analysis of the visual range of orthotropic surfaces of albedo A seen against the sky shows that if for small albedos ($A<0.3$) the visual range is fairly close to that of black bodies ($\approx 4\%$), it departs considerably from the black-body value for light objects ($A>0.3$). In the latter case, moreover, it depends on the azimuth of the object with reference to the sun. This is because as the illumination changes the brightness factor $b=\frac{B}{B_h}$ of such surfaces also changes. In observations against the sun the luminance of objects will be small and that of the sky large, and therefore the brightness factor and visual range of the objects will also be small. In observations in the direction of the sun's rays the relationship between the luminances of object and sky will be reversed.

The situation is further complicated by the fact that real objects have different sizes. For a more complete consideration of the effect of object dimensions, let us turn to expression (14) for the object contrast K, which

we write as follows:

$$K = \frac{K_0}{1 + \frac{1}{b}(p^{-l} - 1)}, \tag{20}$$

where b is the brightness factor of the background or object depending on which is greater.

However, this expression, as we mentioned above, holds only for objects having sufficiently large angular dimensions. For objects with angular dimensions $< 20'$ one must take into consideration the relation (9), which allows for the dependence on the object's dimensions of the limiting contrast K for which the object can still be detected.

Thus if the visual range L of real objects of large angular dimensions can be determined from equation (20) with $K = \varepsilon$ and $l = L$, for objects of small angular dimensions equations (20) and (9) must be solved simultaneously. But in both cases the parameters appearing in these equations must be determined in advance.

In addition to the factors just indicated, the visual range of real objects is affected by their color and contours and, furthermore, by changes in the properties of vision dependent on the illumination level and conditions of observation. This should always be remembered when choosing observation subjects. Thus the principal conditions that determine the visual range of real objects are:

1) meteorological visual range, which characterizes the optical properties of the atmosphere (coefficient α);

2) properties of objects:
   a) true contrast K_0 between object and background,
   b) brightness factor b (larger of b_0 and b_b),
   c) angular dimensions γ of the object and its shape;

3) state of observer's vision:
   a) threshold of brightness contrast of the eye ε,
   b) visual acuity of observer $\left(v = \frac{1}{\delta}\right)$.

If an object for which the visual range is to be determined lies in the field of vision of the observer, by carrying out additional observations one can determine those of the above parameters which are required. In practice, however, one is often faced with the problem of indicating the visual range for specified conditions of observation of a definite object not occurring in the field of vision of the observer. This problem can only be solved approximately by graphical or computational methods. Average values (tabulated) of the parameters for definite typical objects obtained by direct measurement are employed By selecting the values corresponding most closely to the actual conditions of the problem one can give the real visual distance L of the concrete object of interest more or less reliably for various conditions of observation.

We shall not dwell further on these special methods of computation.

In conclusion let us briefly consider the question of nonhorizontal visibility, as it is called, in which the direction of the line of sight to the object is at a certain angle to the horizontal. This may happen, for example, when objects lying at a certain height in the atmosphere (flying aircraft) are observed from the surface, or when objects lying at the surface (aerodrome or landing strip) are sighted from a certain height (flying aircraft); the latter is, of course, especially important from the practical standpoint. This highly complex problem has not been solved exactly to date and despite

a wealth of researches only particular solutions suitable for calculating the visibility of ground objects in the simplest cases are available.

The difficulties encountered in the solution of this problem are due primarily to the fact that neither illumination conditions nor atmospheric transparency remain constant along slanted directions and both often display highly irregular variations. Furthermore, the angular dimensions of objects in the atmosphere are usually small and the objects themselves are not as a rule black. Ground objects are discerned not against a sky background but against the earth's surface, which varies widely in character; therefore the contrast between object and background is generally unknown and moreover strongly dependent on illumination conditions. In view of the complexities involved we will not go into the particular solutions proposed for this important problem but we note that in practical work (aviation) it is usual to employ empirical correlations between the visibility and the meteorological visual range along the horizontal.

§ 6. Visibility of light sources

The determination of the visual range of objects at night or at twilight is a highly complicated affair which depends chiefly on the state of the visual functions of the eye. The nighttime meteorological range which is most important from our point of view can be determined, however, by observing the visual range of artificial light sources (signals): this in itself constitutes a problem of practical importance (visibility of signals, beacons, traffic lights, etc.).

Let us consider the visual range of a single light source. Assume that it lies so far away that it can be treated as a point source. Then its visibility will be determined by its brilliance, i.e., by the illuminance E due to it on the pupil of the eye. This illuminance depends: 1) on the luminous intensity of the light, 2) on the distance l from the light to the observer, and, finally, 3) on the transparency of the atmosphere p. Consequently, it is given by

$$E = \frac{I}{l^2} p^l = \frac{I}{l^2} e^{-\sigma l}. \tag{21}$$

We are assuming that marked air light is absent at night and has no effect. Obviously, the light will be visible as long as E is greater than the threshold illuminance at the eye, i.e., than the minimum illuminance E_0 the eye can perceive. This quantity varies to some extent for different persons and depends strongly on conditions of observation. On the average for calculations it is conventional to take $E_0 = 2 \cdot 10^{-7}$ lux for observations under natural conditions.

Setting the illuminance E equal to the threshold illuminance E_0 in formula (21), we find the limiting distance L beyond which the light is not visible, i.e., the visual range of a light source is given by

$$E_0 = \frac{I}{L^2} e^{-\sigma L}. \tag{22}$$

If the luminous intensity E_0 is known and I is determined from observations, then taking L equal to $2 \cdot 10^{-7}$ lux one can obtain the value of $p = e^{-\sigma L}$,

after which the meteorological visual range can be found from formula (18); nomograms are usually used instead of calculations. It should be noted that considerable technical difficulties are encountered in the organization and performance of such observations on a large scale. The regularities governing the reception of grouped lights are even more complicated and have so far hardly been studied.

A much more complex question and one that is even less explored is that of the visibility of objects illuminated artificially, for instance by searchlights. It can be solved only in individual particular cases The difficulties encountered in its solution stem from the fact that the luminance of an object is determined not only by the luminous intensities of the sources illuminating it but also by their distances from the object and their distribution with reference to the latter. At the same time one must allow for the effect of the haze formed on the sector of the visual cone illuminated by the sources.

We note that in the practical determination of the visual range of light sources, special difficulties arise when one is required to determine their color (e. g., for traffic lights and light signals). In this case the color threshold— the illuminance corresponding to the instant of recognition of color— becomes significant. Many studies show that the color threshold for green light is greater (by a factor of 2–3) than for red light. Therefore red is always easier to distinguish, and moreover is identified almost as soon as it is detected.

§ 7. Visibility in fog and rainfall zones

The pronounced deterioration of visibility in fog and rainfall zones is due to the fact that rain and fog droplets enhance both attenuation (extinction) and scattering of light [sic], thereby reducing contrast.

Turning to the effect of fog on visual range, let us derive the relationship between the liquid-water content and the meteorological visual range, confining ourselves to the case of monodisperse fog. The attenuation of light in fog is determined mainly by the action of the fog particles, since it is so large in comparison with the effect of the air molecules that the latter can be neglected. Therefore one need consider only the attenuation due to the droplets. For the droplet sizes usually observed in fog, according to Mie's theory the coefficient of attenuation due to scattering, which we denote by a, is

$$a = Nk\pi r^2, \tag{23}$$

where N is the number of droplets per unit volume, r is their radius (identical in the case of monodisperse fog), and, finally, k is the scattering function; in our case it can be set equal to 2.

The liquid-water content W can be written as

$$W = N\frac{4}{3}\pi r^3 \rho_d, \tag{24}$$

where ρ_d is the (water) droplet density, equal to unity.

Introducing the value of N from (24) into (23), we obtain the value of a,

after which, making use of the expression for the meteorological visual range

$$S_m = -\frac{\ln c}{\alpha} = \frac{3.91}{\alpha},$$

we obtain the desired relation

$$S_m = 2.61 \frac{r}{W} \rho_d. \qquad (25)$$

From the above one can see that the visual range in fog is inversely proportional to the liquid-water content and directly proportional to the radius of the fog droplets. Thus for the same water content the visual range in a fine-droplet fog will be lower than that in a large-droplet fog.

For polydisperse fog the problem is of course more complicated, since allowance must be made for the size distribution of the droplets. One might mention that the visual range drops to zero in thick fog and decreases to three points ($L<1$ km) even in light fog. Our statements relating to visibility in fog are also valid for visibility in clouds, in which it may drop to 10–20 m (shower clouds) depending on the liquid-water content.

The question of visual range in the rainfall zone and its relationship to the rainfall intensity is complicated by the necessity of considering the size distribution of raindrops; this must be known for a series of points in the line of sight which is nearly impossible in practice. Theoretical consideration of the dependence of the scattering coefficient α upon the intensity I of the rainfall leads to the expression

$$\alpha = MI^b,$$

where M and b are coefficients to be determined; this can be done by calculation from a curve of size distribution of the drops.

The relation $\alpha = 0.25 I^{0.63}$ was obtained from generalized American data. E. A. Polyakova obtained the analogous relation $\alpha = 0.22 I^{0.71}$ (here α is expressed in km^{-1} and I in mm/hr). Calculations and observations show that visibility in the rainfall zone drops 2–3 degrees if the atmosphere were highly transparent before the rain.

Chapter 26

REFRACTION OF LIGHT IN THE ATMOSPHERE

Refraction of light in the atmosphere is the term applied to the bending of the trajectory of a light ray which takes place due to vertical changes in the density of air. Owing to this rays of light are propagated in the atmosphere not rectilinearly but, in general, along a curve. This deflection of the trajectory gives rise to many phenomena, which will be described briefly in the following. We note that the refraction of light is merely a particular case of the refraction experienced by all electromagnetic waves (e. g., radio) and also by sound waves propagating in the atmosphere.

§ 1. Equation of the trajectory of a light ray in the atmosphere

The refractive index n of air depends on its density. For dry air this relationship can be written as

$$n = 1 + a\rho = 1 + A\frac{p}{T}, \qquad (1)$$

where a (respectively A) is a constant which differs for different wavelengths, and $\rho = \frac{pT_0}{p_0T}$ is the relative density of air.

On the average for the region of visible rays $n = 1.000294$. Table 80 lists the values for different wavelengths.

TABLE 80

Values of refractive index of dry air ($t = 0°$, $p = 760$ mm) for different wavelengths (mμ)

Wavelength (mμ)	Refractive index
400	1.0002983
500	1.0002943
550	1.0002931
600	1.0002922
700	1.0002910
800	1.0002902

Expression (1) shows that the refractive index depends on air pressure and temperature, the dependence on the latter being decisive. Atmospheric

humidity also has a certain effect but in the visible spectrum this is so small that it need not be considered.

Let us divide the atmosphere into a series of thin concentric layers, inside each of which we will take the density to be constant (Figure 166). From the above statements it follows that the refractive index decreases from layer to layer with height. A ray moving from one layer into the adjacent, underlying layer will be refracted towards the normal, i.e., will be

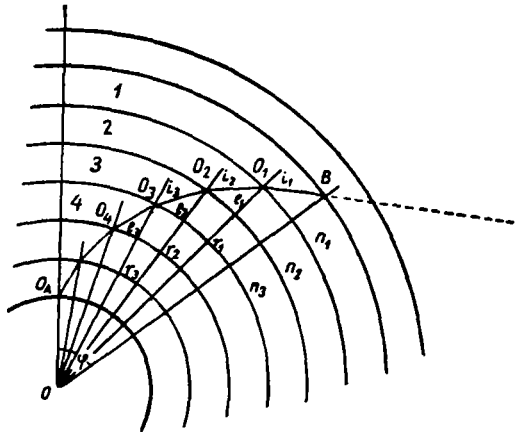


FIGURE 166

deflected towards the earth's surface. We shall obtain the equation for the trajectory of the ray in this very simple case. Let us consider the consecutive layers 1, 2, 3, etc., and denote the refractive indexes for these by n_1, n_2, n_3, etc. Further, let the angles of incidence at the boundary of transition from layer 1 to layer 2, layer 2 to layer 3, etc., be i_1, i_2, i_3 and the angles of refraction correspondingly e_1, e_2, e_3, ... Then, obviously, one can write

$$\frac{\sin i_1}{\sin e_2} = \frac{n_2}{n_1}, \quad \frac{\sin i_1}{\sin e_3} = \frac{n_3}{n_2} \text{ etc.}, \tag{2}$$

on the other hand, from the triangles O_1OO_2, O_2OO_3, etc., we have

$$\frac{\sin e_1}{\sin i_2} = \frac{r_2}{r_1}, \quad \frac{\sin e_2}{\sin i_3} = \frac{r_3}{r_2} \text{ etc.}, \tag{3}$$

where r_1, r_2, r_3, etc., are radius-vectors.

Eliminating the angles e_1, e_2, etc., from (2) and (3), we obtain

$$\frac{\sin i_1}{\sin i_2} = \frac{n_2 r_2}{n_1 r_1} \quad \text{or} \quad n_1 r_1 \sin i_1 = n_2 r_2 \sin i_2,$$

$$\frac{\sin i_2}{\sin i_3} = \frac{n_3 r_3}{n_2 r_2} \quad \text{or} \quad n_2 r_2 \sin i_2 = n_3 r_3 \sin i_3.$$

In the general case

$$nr \sin i = \text{const} = A, \tag{4}$$

which gives the equation for the trajectory of a ray.

Equation (4) can also be written in the polar coordinate system. Placing the origin at the earth's center, one can easily show with the aid of Figure 167 that the equation for the trajectory of a ray can be written as

$$\varphi = \int_{r_1}^{r_2} \frac{\sin i}{\sqrt{1-\sin^2 i}} \frac{dr}{r}, \tag{4'}$$

(the notation is made clear by the figure).

Since from (4) $\sin i = \frac{A}{nr}$,

$$\varphi = \int_{r_1}^{r_2} \frac{A}{\sqrt{n^2r^2 - A^2}} \frac{dr}{r}. \tag{4''}$$

Let us consider in greater detail the question of the vertical variation of n and curvature of the trajectory of the light ray. Recalling that $\rho = \frac{pT_0}{p_0 T}$, we differentiate (1) with respect to the height

$$\frac{dn}{dh} = a\frac{d\rho}{dh} = a\frac{T_0}{p_0}\left[\frac{1}{T}\frac{dp}{dh} - \frac{p}{T^2}\frac{dT}{dh}\right]. \tag{5}$$

In view of the hydrostatic equation $\frac{dp}{dh} = -\frac{pg}{RT}$ we obtain

$$\frac{dn}{dh} = -a\frac{T_0 p}{p_0 T^2}\left[\frac{g}{R} + \frac{dT}{dh}\right]. \tag{6}$$

Here the constant $\frac{g}{R} = 0.0341°/\text{m}$ is known as the autoconvective temperature gradient (temperature gradient in a homogeneous atmosphere).

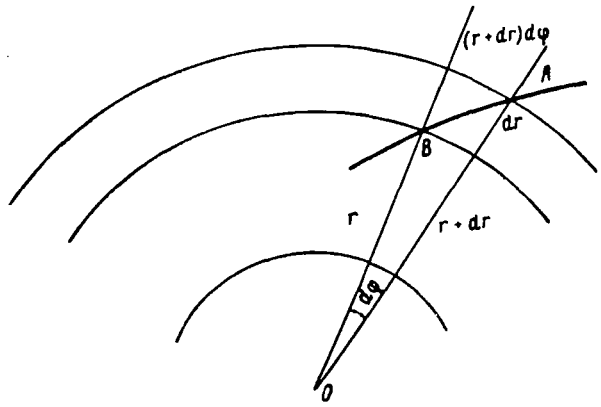


FIGURE 167

Indeed, (6) shows that if $\frac{dT}{dh} = -0.0341$ then $\frac{dn}{dh} = 0$ and hence n does not change with height.

If we confine ourselves to the lowermost layers then, taking $p = p_0$ and $T = T_0 = 273°\text{K}$ we have the following values for $\frac{dn}{dh}$ as a function of $\frac{dT}{dh}$ °/m

(from 6)

$\frac{dT}{dh}$	°/100 m	−1.01	−0.005	0.000	+0.1	+0.1
$\frac{dn}{dh}$		$-2.58\cdot10^{-8}$	$-3.11\cdot10^{-8}$	$-3.65\cdot10^{-8}$	$-14.4\cdot10^{-8}$	$-111\cdot10^{-8}$

i.e., the vertical changes in the refractive index are usually small.

Let us denote the radius of curvature of the trajectory of a ray by R_s. Then if the element of trajectory is ds and the change in the angle of incidence di we can write the following expression for the curvature of the ray:

$$\frac{1}{R_s} = \frac{di}{ds}. \tag{7}$$

But $dh = ds\cos i$, whence $ds = \frac{dh}{\cos i}$, and it will be shown later (cf. equation (10)) that $di = -\operatorname{tg} i \frac{dn}{n}$. Therefore, recalling (6) and (1), we find

$$\frac{1}{R_s} = -\frac{\sin i}{n}\frac{dn}{dh} = \frac{a\sin i}{1+a\rho}\frac{T_0 p}{p_0 T^2}\left[\frac{g}{R} + \frac{dT}{dh}\right]. \tag{8}$$

At normal pressure ($p = p_0 = 760$ mm Hg) and temperature ($T = T_0 = 273$°K) we obtain

$$\frac{1}{R_s'} = 1.07\cdot10^{-6}\left[\frac{g}{R} + \frac{dT}{dh}\right]\sin i \text{ km}^{-1}, \tag{9}$$

which for $i = 90°$, i.e., horizontal rays, gives

$$\frac{1}{R_s'} = 1.07\cdot10^{-6}\left[\frac{g}{R} + \frac{dT}{dh}\right] \text{km}^{-1}, \tag{9'}$$

this shows that the radius of curvature R_s of a ray depends on the vertical temperature gradient $\frac{dT}{dh}$.

Table 81 gives the values of R_s' for different $\frac{dT}{dh}$.

TABLE 81

$\frac{dT}{dh}$ (°/100 m)	−3.4	±1.0	±0.5	0.0	+6.9	+11.4
R_s' (km)	∞	39 000	31 800	27 700	9100	6370

The table shows that a ray for which $i = 90°$ at the given point (in the particular case, tangent to the earth's surface) will propagate rectilinearly ($R_s' = \infty$) only if the temperature drops with height at a rate of 3.4° per 100 m. For a higher lapse rate the ray will be convex towards the earth's surface and will travel away from the surface, since the earth's radius is 6370 km. For a lapse rate smaller than 3.4°/100 m and when the temperature increases with height the ray will be concave towards the surface; only for a temperature rise of 11.4°/100 m or more will the radius of curvature be equal to or smaller than the earth's radius. Consequently, a ray emerging from any point on a certain surface corresponding to a definite refractive index (in particular, the earth's surface) will, under certain conditions, be concave to the earth's surface and be able to return to another point on this surface.

§ 2. Astronomical and terrestrial refraction

Owing to the bending of the trajectories of rays, more or less distant objects are always seen along directions slightly different from those in which they actually occur, since we project an object we observe in the direction of the tangent to the trajectory of the ray at the point where it enters the eye. The angle γ between the tangents at the initial point B and the final point A of the ray's path is called the total angle of refraction or sometimes simply the total refraction. Obviously, it gives the change in the ray's direction in the atmosphere (Figure 168). The angle α between the direction from the eye to the true position of the sighted point and the direction in which we see this point, i.e., the direction of the tangent to the trajectory of the ray at its terminal point (the eye), is called the angle of refraction or simply the refraction. If the sighted point lies outside the atmosphere (celestial body) the refraction is described as astronomical refraction; if the sighted point lies within the atmosphere, the refraction is termed terrestrial refraction.

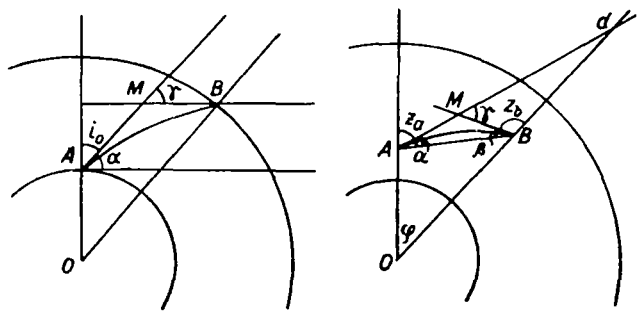


FIGURE 168

Let us obtain the general expression for the total angle of refraction. Returning to Figure 166, we take any two adjacent layers (1 and 2, say); let their refractive indices be $n_1=n$ and $n_2=n+dn$. Further, denoting the angle of incidence of the ray at the surface of discontinuity by $i_1=i$ and the angle of refraction by $e_1=i_2=i+di$, we write

$$\frac{\sin i}{\sin (i+di)}=\frac{n+dn}{n}.$$

Confining ourselves to first-order infinitesimals, we have

$$\sin i=\frac{n+dn}{n}\sin (i+di)=\sin i+\cos i\,di+\sin i\frac{dn}{n},$$

whence

$$di=-\operatorname{tg} i\frac{dn}{n}. \tag{10}$$

But di is the change in the direction of the ray on transition from the first layer to the next, i.e., irrespective of the sign it is the differential

of the refraction $d\gamma$, and therefore

$$\gamma = \int_{n_A}^{n_B} \operatorname{tg} i \frac{dn}{n}, \tag{11}$$

where n_A and n_B are the refractive indices at the initial and terminal points of the ray's path.

In view of the fact that

$$\operatorname{tg} i = \frac{\sin i}{\sqrt{1-\sin^2 i}} = \frac{n_0 r_0 \sin i_0}{\sqrt{n^2 r^2 - n_0^2 r_0^2 \sin^2 i_0}},$$

we obtain

$$\gamma = \int_{n_A}^{n_B} \frac{n_0 r_0 \sin i_0}{\sqrt{n^2 r^2 - n_0^2 r_0^2 \sin^2 i_0}}$$

Since n depends on temperature and pressure, in order to make use of (11) it is necessary to know the law governing the variation of T and p with height. This law will be different in every individual case and in practical work it is usual to employ tables calculated for a certain average state of the atmosphere.

For an approximate calculation of the astronomical refraction one can use the formula

$$\gamma = 57'' \operatorname{tg} z', \tag{12}$$

or

$$\gamma = 57'' \operatorname{tg} z' - \frac{0.55 \sin z'}{\cos^2 z'}, \tag{12'}$$

where z' is the apparent zenith distance of the body, which is always smaller than its true zenith distance z; $z' = z$ only for $z = 90°$.

The above formulas give sufficient accuracy for many practical purposes, the first for zenith distances not exceeding 80° and the second for zenith distances not exceeding 89°

Table 82 lists the refractions for $p = 750$ mm and $t = 8.5°$.

TABLE 82

Values of refraction γ

z' (deg)	90	89.7	89.3	89.0	88.0	85.0	70	60	50	40	30	20	10	0
γ (min)	34.9	30.9	27.4	24.4	18.2	9.8	2.6	1.7	1.2	0.8	0.6	0.4	0.2	0

As one can see, the refraction is particularly large in the region of large zenith distances, in which it varies very sharply with z'.

In the case of terrestrial refraction (Figure 168) the total refraction will be equal to the sum of the angles of refraction α and β observed when looking at the point B from the point A and vice versa, i.e., $\gamma = \alpha + \beta$. In the general case α is not equal to β but in practice one usually takes $\alpha = \beta$ and hence $\alpha = \frac{1}{2}\gamma$, since for the comparatively limited distances between points A and B situated at not very different heights the trajectory of the ray is a nearly regular arc of circle; therefore,

$$\alpha = \beta = \frac{1}{2}\gamma = 90° - \frac{z_A - z_B}{1} + \frac{\varphi}{2}. \tag{13}$$

Since φ is known and z_A and z_B are determined from observations, one can easily find the terrestrial refraction α. To give the reader an idea of the magnitudes involved, we list its mean values for different distances D from the sighted point:

D (km)	1	4	8	12	16	20
α (")	2	8	17	25	34	42

For an exact determination of the refraction, of course, more complicated procedures must be adopted.

§ 3. Phenomena due to astronomical refraction

The effect of astronomical refraction is felt when one observes the celestial bodies (sun, moon, stars, etc.). Owing to astronomical refraction the observed zenith distances of heavenly bodies seem smaller than the true ones. This effect is particularly strong for bodies lying near the horizon, and therefore we see the sun and other celestial bodies some time after they have set and also a little before they rise. As a result the day is 8–12 min longer in middle latitudes and still longer in polar regions (where the angle between the horizon and the plane of great circle along which the sun moves is very small); beyond the Arctic Circle the length of the polar day can increase by several days owing to refraction.

The next phenomenon due to refraction is the change in the shape of the solar and lunar disks near the horizon. Since the lower rim of the sun near the horizon is raised by 35' owing to refraction while its lower rim is raised by only 28', the vertical diameter appears to be shorter by 7' and the setting or rising sun often seems flattened along the vertical. A similar deformation is observed for the lunar disk.

Since the refractive index is different for rays of different wavelengths, the effect of astronomical refraction will also vary and a beam of white light entering the atmosphere will be dispersed into its component colors as it travels through it. In the case of yellow ($\lambda = 500\,m\mu$) and violet ($\lambda = 400\,m\mu$) rays the dispersion for different zenith distances of the celestial body is:

z (°)	50	65	85	89	90
Dispersion (")	1	4	11	26	38

i. e., it is particularly large at the horizon. Accordingly, a luminous point (star) might be drawn out in the form of a rainbow band (spectrum) near the horizon— as can be observed in reality.

An analogous phenomenon should take place for the sun but because the sun's apparent diameter is 32' and the dispersion of light amounts to 38", a large portion of the color images overlap and produce a white color, and only the extreme colors remain pure. Further, since the attenuation of light increases very rapidly with decreasing wavelength, usually only the longer rays remain in the spectrum at the instant of sunset. However, under especially favorable conditions of atmospheric transmission green rays and even blue rays (admittedly very seldom) may remain. Thus at

sunset, just at the instant when the topmost point of the sun's disk dips below the horizon, under favorable conditions one may observe a rapid succession of colors in the narrow light band still remaining, from red-yellow to green and sometimes even to blue. This phenomenon of the "green flash" is observed only under particularly favorable conditions of transmission and usually lasts a very short time— from fractions of a second at low latitudes to a few seconds at high latitudes.

Also linked to the refraction of light is the twinkling of stars at elevations above the horizon (not exceeding 50°)—the stars appear to flicker and change brightness as well as, occasionally, color. This phenomenon is due to extremely rapid and irregular changes in the refractive index in the air layers crossed by the rays on their way to our eyes.

The phenomenon of shadow bands, a term which designates the alternating light bands often seen moving in a certain direction over a smooth landscape surface, can be explained in a similar way. It is due to light rays crossing small formations of more or less dense air moving in the atmosphere in some direction. A similar explanation holds for the scintillation of distant objects often observed on hot days, when intensive mixing of air prevails in the strongly overheated lower layers of the atmosphere.

§ 4. Lifting and lowering of horizon. Anomalous refraction

Terrestrial refraction also gives rise to a series of phenomena, notably the lifting and simultaneous broadening (and also lowering and narrowing) of the horizon. Let us suppose (Figure 169) that the eyes of the observer

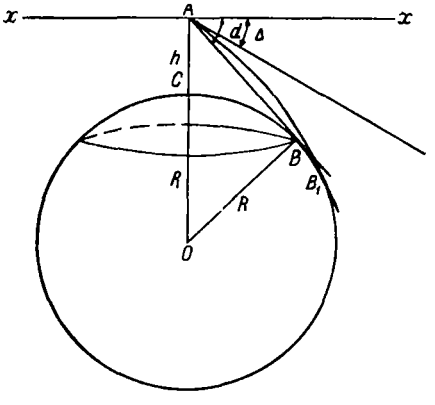


FIGURE 169

are at the point A at the height h above the smooth earth, here assumed to be a sphere of radius R, and that refraction is absent. The line of the apparent horizon will be the small circle on the sphere representing the geometric locus of the points of contact of straight lines drawn tangent to the earth's surface from the point A. The distance of the horizon D, represented on the figure by the segment AB, is easily found in this case

from the triangle AOB

$$D = \sqrt{(R+h)^2 - R^2}.$$

In view of the smallness of h compared with R, the above can be written as

$$D = \sqrt{2Rh} = 3.57\sqrt{h} \text{ km} \tag{14}$$

(here D is measured in kilometers and h in meters).

The apparent horizon (surface of cone with generatrix AB) subtends a certain angle d (so-called geodetic depression of horizon) with the true horizon (plane xx). This angle can be found from the equation

$$\operatorname{tg} d = \frac{AB}{OB} = \sqrt{\frac{2h}{R}}, \tag{15}$$

which for small d gives

$$d^2 = \frac{2h}{R}. \tag{16}$$

Owing to refraction, however, the line of sight is actually a curve rather than a straight line and the ray moves along a curve tangent to the earth's surface at the point B_1, which the eye perceives in the direction forming the angle Δ with the true horizon (observed depression).

The general equation for the trajectory of the ray AB_1 in this case can be written as

$$Rn_0 \sin i_0 = (R+h)\, n \sin i,$$

where n_0 and n are the refractive indices at the earth's surface (point B_1) and at the height h (point A); $i_0 = 90°$ and i are the angles of incidence at the points B_1 and A respectively.

From this

$$\sin i = \frac{Rn_0}{(R+h)\, n}. \tag{17}$$

But $i = 90°$ and $n = n_0 + \Delta n$; therefore from the above we obtain

$$\cos \Delta = \frac{Rn_0}{(R+h)\,(n_0 + \Delta n)} \tag{18}$$

and

$$\operatorname{tg}^2 \Delta = \frac{\sin^2 \Delta}{\cos^2 \Delta} = \frac{(R+h)^2}{R^2} \frac{(n_0+\Delta n)^2}{n^2} - 1. \tag{19}$$

Owing to the smallness of the angle Δ equation (19) can be written in the form

$$\Delta^2 = \frac{(R+h)^2 (n_0+\Delta n)^2}{R^2 n^2} - 1. \tag{20}$$

From here, disregarding quantities of the second order, we obtain

$$\Delta^2 = \frac{2h}{R} + \frac{2\Delta n}{n_0}, \tag{21}$$

or, bearing (16) in mind and recalling that n_0 is close to unity,

$$\Delta^2 = d^2 + 2\Delta n = d^2 + 2(n - n_0). \tag{22}$$

According to the above, as a result of refraction the apparent horizon will be slightly raised $(\Delta < d)$ if $n_0 > n$ and depressed $(\Delta > d)$ if $n_0 < n$.

The greater the difference between the refractive indices at the surface and at the point A (eye of the observer), the greater this lifting.

The lifting of the line of the horizon is accompanied by an increase in its distance, which is given by

$$D_1 = \sqrt{2Rh}\left(1 + \frac{kR}{2}\right) = D\left(1 + \frac{kR}{2}\right), \tag{23}$$

where

$$k = -\frac{dn}{dh}.$$

If we assume that at normal pressure and temperature $k = 2.58 \cdot 10^{-8}$, then $kR = 0.164$ and

$$D_1 = D(1 + 0.08) = 1.08D, \tag{24}$$

which corresponds closely to reality.

Table 83 indicates the values of D and D_1 as calculated from (24) for different heights h of the point A, and also the values of d and Δ.

TABLE 83

Lifting and broadening of horizon owing to refraction

Height of observation (m)	1	10	100	1000
Geodetic depression d	2'	6'	21'	1°5'
Observed depression Δ	2'	5'	19'	1°
Geometric distance of horizon D(km)	3,5	11	36	113
Observed distance of horizon D_1 (km)	4,0	12	38	122

Formulas (22) and (23) show that the true depression will coincide with the geodetic depression and the visual range of the horizon D_1 with the geometric visual range D only when $n = n_0$.

If $n > n_0$, as when the density increases with height, the trajectory of the ray will be convex towards the earth's surface and the horizon will seem depressed and narrow compared with its geometric position. Since in the ordinary state of the atmosphere $n_0 > n$, $\Delta < d$ and $D_1 > D$, i.e., there will always be a certain lifting and broadening of the horizon. But this mean lifting of the horizon is small and we are accustomed to it. When the state of the atmosphere (distribution of density) changes, however, this lifting can sometimes change fairly significantly. We can then see much farther, e.g., objects usually hidden below the horizon, and the edges of the horizon and objects seem to be slightly raised this is the so-called anomalous positive refraction. When $n - n_0$ becomes perceptibly smaller than the mean and especially when its sign changes ($n_0 < n$), the horizon seems depressed and narrow and we may be unable to see objects on the horizon that are visible under ordinary conditions; this is the so-called anomalous negative refraction.

Instances of anomalous refraction, both positive and negative, are fairly common.

§ 5. Mirage

Layers of sharply contrasting density may be formed in certain special states of the atmosphere, leading to drastic changes in density and refractive index at the surface of discontinuity and causing the many and various phenomena known as mirages.

Instead of the single image of the distant object that one usually sees, in mirages there are several images lying one on top of the other, some of which may even be inverted. This feature of the phenomenon inidcates that the rays from the object take different paths through the atmosphere; we may perceive several rays from each point, projecting the object along the tangents to each of these rays. We note that sometimes the object itself may lie outside our field of vision (below the horizon). Obviously, this can happen only when there is total reflection of the incident rays at the surface of discontinuity between the layers. Such reflection can be observed in the atmosphere only when rays moving into one with a lower refractive index are incident at an angle greater than the angle of total internal reflection; in other words, the angle of incidence of the rays at the surface of discontinuity is greater than the value of α' satisfying the relation $\sin\alpha' = \frac{n_2}{n_1}$, where n_2 and n_1 are the indices of refraction in the denser and lighter layer, respectively. Since the difference between n_2 and n_1 is small, the limiting angle α' will be very slightly different from 90°; for example, even for the density difference due to a temperature jump of about 10° at the discontinuity, the angle α' is greater than 89.5°.

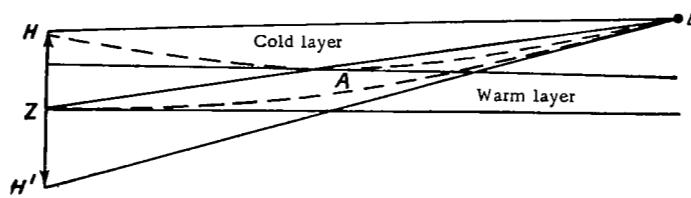


FIGURE 170. Mechanism of inferior mirage

One speaks of an inferior mirage if the reflected image lies below the object and of a superior mirage if the reflection lies above the object.

Schematically the formation of an inferior mirage can be described as follows (Figure 170). Suppose that vigorous warming of the surface by solar rays leads to the formation of a thin layer of strongly overheated air near the ground, and that this layer, in which the density and refractive index are small, is overlain by layers of greater density. Such a state would of course be temporary and unstable and could easily be disrupted, for which reason it is usually observed only in the absence of wind. In this case if one looks at any object ZH from the point B, then the eye (point B) may receive two rays from any point of the latter (H, say): a ray which travels entirely through the upper cooler and denser air (ray HB) and is only slightly bent, and a ray which, travelling downwards, is reflected at the point A on the surface of discontinuity and arrives along the path HAB. Projecting the point H along the tangent to this ray, we see its image at the point H'. The same thing can happen for the other points of the object ZH, the image of which will be seen below the true one and will moreover be inverted.

Inverted mirages often take the form of a sector of the horizon sky. The image of it that the eye sees is a bluish surface like water obscuring part of the landscape; distant objects therefore appear to be floating over a surface of water (actually the reflection of part of the sky) lying in their foreground. This phenomenon is frequently observed during summer calms on the steppes, in deserts, over the surfaces of asphalted roads, etc. The peculiar changes which sometimes occur in the shape of the sun or moon when they set are also due to a certain extent to refraction.

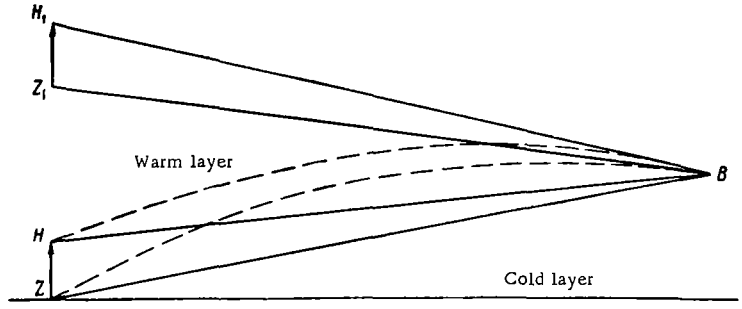


FIGURE 171. Mechanism of superior mirage

Superior mirages are formed in the converse situation when the earth's surface and adjoining layer of air are strongly cooled (usually in winter). If the transition from this cool layer to the overlying warmer layers is sufficiently sharp, conditions will favor the reflection of rays issuing from

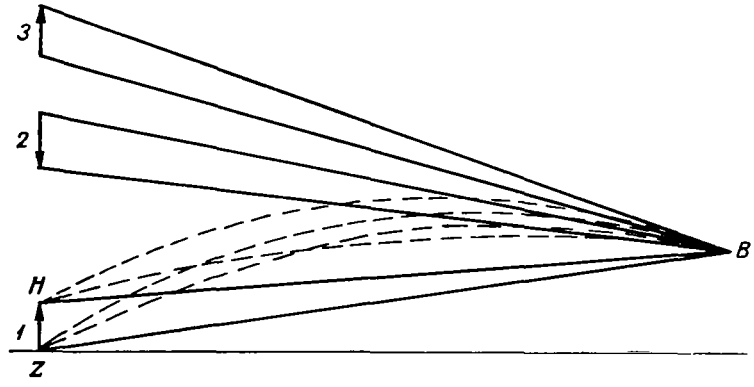


FIGURE 172. Mechanism of complex mirage

from sighted objects (Figure 171). Moreover, several images of the object, both upright and inverted, may be formed, depending on how many rays enter the eye and by what paths they travel in the atmosphere. This is illustrated by Figure 172, which shows the conditions that can lead to the formation of two images besides the true one (1), an upright image (3) and and inverted one (2).

Obviously, the diversity of conditions of vertical density distribution in the atmosphere can lead to a wide variety of images perceptible to the eye, some of them completely unexpected. Recalling that the image seen is not stable and is in a state of perpetual motion, one can understand the origin of the truly fantastic visions called fata morgana— images (embellished by the imagination) of fairytale cities, gardens and the like. We shall not discuss example of mirages described in the extensive literature on this subject or known from literary works. Equally, we will not consider theoretical studies in which the authors have attempted to express the mechanism of mirages in mathematical form, in view of their specialized nature.

Chapter 27

OPTICAL PHENOMENA DUE TO THE PRESENCE OF WATER DROPLETS AND ICE CRYSTALS IN THE ATMOSPHERE

When rays of light travel across clouds and rainfall zones they give rise to special optical phenomena — rainbows, coronas, halos, etc. — most of which are familiar to everyone. All these phenomena can be explained by the refraction, reflection and diffraction of light on cloud elements. The present chapter is devoted to a brief consideration of these.

§ 1. Rainbows

Rainbows appear in rainfall zones, usually against a background of clouds in the portion of sky away from the sun (the center of the arc lies at the antisolar point). The radius of the rainbow is 42°, but a second rainbow concentric with the first and having a radius of about 52° can often be observed. In the first rainbow the red ring is on the outside and the violet one on the inside while in the second the ordering of the colors is reversed. The intensity and width of the individual colors vary and depend, as the theory shows, on the dimensions of the water droplets responsible for the appearance of the rainbow. Sometimes secondary color arcs (up to six) may be visible on the inside of the first rainbow and outside of the second, usually displaying an alternation of green and pink In a few cases where the raindrops are very small a white rainbow or fogbow is observed.

Rainbows seen in moonlight generally look like white rainbows. In contradistinction with the latter, however, the white color of moonlit rainbows is due chiefly to the properties of the retina of the eye (at twilight the eye distinguishes colors poorly) and not to the dimensions of the raindrops.

The phenomenon of the rainbow is caused by the reflection and refraction of the sun s rays in water drops. The earliest (though not altogether complete) explanation of the rainbow was given already in 1637 by Descartes; it was later completed by Newton.

If one considers a monochromatic ray falling on a drop and subsequently emerging from it, one can determine the change which takes place in its direction in this process. Indeed, take a ray of a certain wavelength λ incident on a drop at the angle i. On penetrating into the drop it is refracted; we will denote the angle of refraction by e. One can easily show from Figure 173a that on entry into the drop the direction of the ray changes by the angle $i - e$, in each reflection inside the drop the ray is rotated by $\pi - 2e$, and on emergence from the drop the angle of deviation is again $i - e$.

Consequently, if the ray is reflected D times inside the drop, the angle by which it will be bent from its original direction can be expressed as

$$D=2(i-e)+k(\pi-2e), \tag{1}$$

or

$$D=k\pi+2[i-(k+1)e]. \tag{1'}$$

Let us find the conditions for which D is a minimum. We take the derivative

$$\frac{dD}{di}=2\left[1-(k+1)\frac{de}{di}\right]. \tag{2}$$

Since $\frac{\sin i}{\sin e}=n$, where n is the refractive index, we have

$$\frac{de}{di}=\frac{\cos i}{n\cos e},$$

and instead of (2) we can write

$$\frac{dD}{di}=2\left[1-(k+1)\frac{\cos i}{n\cos e}\right]. \tag{3}$$

Setting the derivative equal to zero and denoting i and e for this case by I and E, we obtain from the above

$$(k+1)\cos I=n\cos E, \tag{4}$$

whence it is easy to find that

$$\cos I=\sqrt{\frac{n^2-1}{(k+1)^2-1}} \tag{5}$$

and

$$\cos E=\frac{k+1}{n}\cos I. \tag{5'}$$

From (1'), which takes on the form

$$D_k=k\pi+2[I-(k+1)E], \tag{6}$$

we obtain the value of D_k for any k ($k=1, 2, 3,$ etc.).

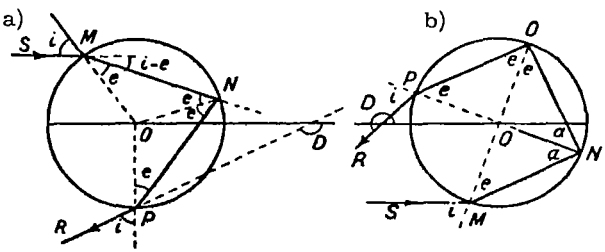


FIGURE 173. Path of ray in water drop

Results of calculations as above for the first and second rainbow ($k=1, 2$) are given in Table 84 for several wavelengths. The luminance maxima producing the image of the rainbow should be observed at the indicated distances.

Thus parallel rays falling on a drop will be divergent when they emerge from it. But from what we said it is clear that, for example, red rays falling at angles close to 59°29' will be the least deflected and will diverge the least; they will issue in a beam deflected from the initial direction by the angle $D_k = 137°42'$ and will display maximum intensity at a distance of 42°18' from the antisolar point. Obviously, according to this theory no red ray will arrive at points located at distances smaller than 42°18' from the antisolar point. All rays of smaller wavelength and correspondingly higher refringibility will have larger values for the angle of minimum deviation; for example, for the violet rays we have $D_k = 139°24'$, and they give maximum intensity at 40°36' from the antisolar point.

TABLE 84

Color	Refractive index n	Angle	First rainbow ($k = 1$)	Second rainbow ($k = 2$)
Violet ($\lambda = 404.7$ mμ)	1.3435	I	58°48'	71°30'
		E	39 33	44 54
		$180° - D_k$	40 36	53 36
Green ($\lambda = 546.1$ mμ)	1.3352	I	59 17	71 46
		E	40 5	45 19
		$180° - D_k$	41 46	51 38
Yellow ($\lambda = 577.9$ mμ)	1.3341	I	59 21	71 48
		E	40 10	45 23
		$180° - D_k$	41 58	51 18
Red ($\lambda = 656.3$ mμ)	1.3318	I	59 29	71 53
		E	40 19	45 31
		$180° - D_k$	42 18	50 40

Thus if the sun were a point the width of the rainbow would be $42°18' - 40°36' = 1°42'$; but since the diameter of the sun is 32' the width of a rainbow is $1°42' + 32' = 2°14'$ and all colors are pure with the exception of the marginal red.

Applying analogous arguments to the case of double reflection of a ray in a droplet (Figure 173b), we find that the width of the rainbow in this case amounts to $53°36' - 50°40' + 32' = 3°34'$, and the ordering of the colors is reversed.

Triple reflection could lead to the formation of a rainbow on the solar side (at a distance of 44° from the sun), but owing to its low intensity against the brightly lighted sky such a rainbow is never observed.

This elementary theory makes it possible to determine correctly the paths of individual rays in the direction of the least deflected ray [Descartes ray], but as it does not take the wave aspects of the propagation of light into account it fails to explain a number of details and is thus incomplete.

Developing the theory further, several writers considered the character of the changes experienced by the surface of a plane wave when passing through a drop. The rigorous theory of this question (admittedly only for monochromatic light) was given by Airy. He showed that the initially plane wave surface becomes bent on emerging from the drop; it is bent in opposite

directions on either side of the least deflected ray and therefore the rays diverge on one side and converge on the other.

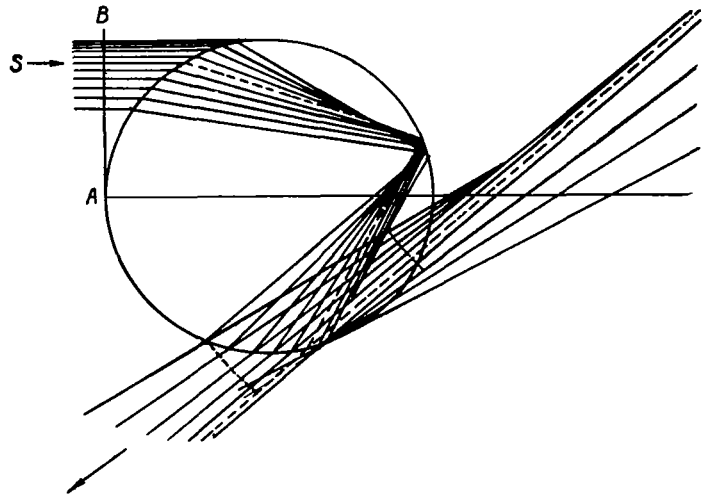


FIGURE 174. Path of rays in drop

The path of the rays and corresponding deformation of the wave surface are shown in Figure 174, where the dotted line represents the path of the least deflected ray. Only those rays that diverge very slightly (i. e., lie

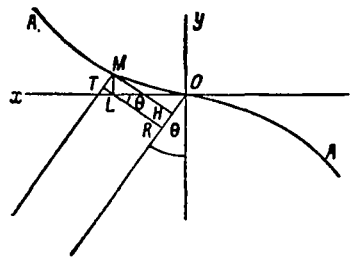


FIGURE 175

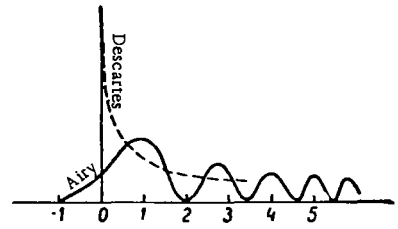


FIGURE 176. Distribution of luminous intensity in rainbow

close to the least deflected ray on either side of it) reach the eye. The small part of the deflected wave surface that participates in this process and gives rise to the rainbow (its active part) is represented in Figure 175. It can be shown that this line is the involute of a caustic surface. Considering the luminous image produced in the eye by this active part of the wave surface, Airy showed that the luminous intensity is expressed as

$$A^2 = M^2 f^2(z). \tag{7}$$

Her A is the amplitude of oscillation, M is a certain factor which

depends on the drop size and wavelength of the rays under consideration, and

$$f(z) = \int_0^\infty \cos \frac{\pi}{2} (w^3 - zw)\, dw$$

is the so-called Airy integral, where $z = \frac{4}{\lambda} \theta \sqrt[3]{\frac{2a^3\lambda}{4h}}$ and $w = x \sqrt[3]{\frac{4h\cos\theta}{3a^2\lambda}}$ (x being the distance of the active point of the involute from the least deflected ray).

For given drop size (radius a) and wavelength λ, z and w are determined by the angle θ between the rays under consideration and the direction of the least deflected ray (Figure 175), and by the quantity $h = \frac{k^2 - 1 \sin I}{k^2 \cos^2 I}$, which depends on the number of reflections in the drop k and on the refractive index.

By calculation of the highly complex integral Airy obtained the luminous intensity distribution shown graphically in Figure 176 (origin coincides with position of least deflected ray). The intensity distribution according to Descartes' theory is also shown (dotted line).

Extending Airy's researches, Pertner considered the intensity distribution of a series of wavelengths for different drop sizes and was able to determine the rainbow form and color distribution that were to be expected according to drop size. It was found that the larger the drops, the stronger the individual colors and the narrower the rainbow; conversely, for very small drop sizes the rainbow would be broad and nearly white with faintly colored edges; this kind of rainbow is sometimes seen in fog. It was also established that for large drops the secondary rainbows are continuous with the violet edge of the main rainbow and though not broad display fairly strong colors.

It has lately been shown that the general theory of the rainbow can be developed from the electromagnetic theory of light by considering scattering on large particles; however, we will not dwell on this specialized question.

§ 2. The halo

When a thin cloud composed of ice crystals lies between the sun (or moon) and the observer's eye, refraction and reflection of the rays crossing the cloud give rise to the whole series of luminous phenomena known as halos. While halos assume many different forms, certain major types are observed most frequently; these are illustrated schematically by Figure 177. Among them halos in the form of a 22°-ring are particularly common.

Halo phenomena can be divided into two groups according to coloring: 1) halos faintly colored in the rainbow hues, in which the red color always occurs in the portion closest to the sun, examples being rings of 22 and 46° radius, mocksuns [sun-dogs, parhelia] of 22° and certain other displays; and 2) colorless (white) halos, for example horizontal circles, pillars, etc. This certainly indicates that colored halos are due to the refraction of rays in ice crystals while colorless halos arise from the reflection of rays by crystal faces.

To understand halo phenomena one should recall that ice crystals belong to the hexagonal system and can assume a wide variety of shapes by combining the different forms of this system. However, halos appear only when certain crystal forms are present. The principal forms capable of giving rise to halso are: 1) crystals in which the principal axis is strongly developed in comparison with the secondary axes, i.e., hexagonal prisms, sometimes with hexagonal pyramids (mostly truncated) added to one or both bases; 2) crystals with a very weakly developed principal axis, i.e., hexagonal plates. All other forms can be regarded as combinations of the above two.

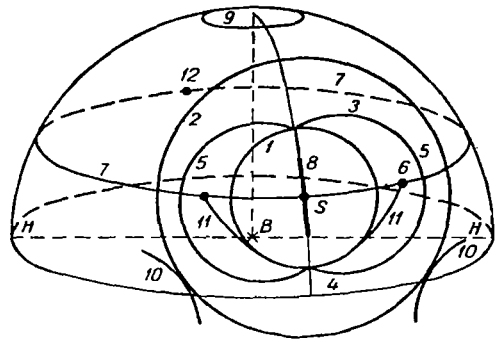


FIGURE 177. Principal halo forms

1–small ring (22°); 2–large ring (46°); 3–upper tangential arc to 22° circle; 4–lower tangential arc to 22° circle; 5–circumscribed halo; 6–mocksun [parhelion]; 7–horizontal circle; 8–pillars; 9–circumzenithal arc; 10–lateral tengential arcs to 46° circle; 11–Lowitz arc; 12–anthelion.

When falling in the atmosphere the crystals tend to orient themselves so as to encounter maximum resistance from the air.

In the theory of halos the orientation of the principal crystal axis must be taken into account. One should consider whether the predominant number of crystals in the cloud have vertically, horizontally or randomly (all possible directions) oriented principal axes. In the first two cases we have rigorously oriented halos, the form of which changes with changing solar elevation; in the last case the halos assume the form of circles with the sun as center.

To explain the formation of the colored halo forms, let us consider the path of a ray in an ice prism in the principal section of the prism. We note in advance that a ray entering a prism across a certain face can emerge from the prism only through faces forming an angle not greater than 99°32' with the face of entrance. Indeed, consider a ray incident on the face ab at the angle i (Figure 178). After being refracted by the angle e, it emerges through the face cd, forming the angle of incidence e' and the angle of emergence i'. Since the refractive index of ice is $n = \frac{\sin i'}{\sin e'} = 1.31$, for the limiting case where $i' = 90°$ (and therefore the ray can no

longer emerge from the prism) we find that $e' = 49°46'$ and $e + e' = A$, where A is the refracting angle of the prism. Owing to the fact that e and e' are not greater than 49°46', A cannot be greater than 99°32'. Thus from our point of view only those faces which form angles smaller than 99°32' with each other are significant; the most interesting of these are 60 and 90°.

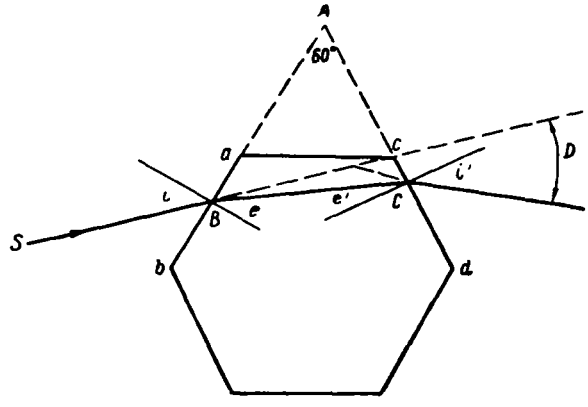


FIGURE 178. Path of ray in ice prism

Returning to Figure 178, let us determine the angle D by which the ray is deflected from its original direction; it is given by

$$D = i - e + i' - e' = i + i' - A. \tag{8}$$

Let us find the minimum value of $D = D_{\min}$. We take the derivative of i with respect to D and set it equal to 0

$$\frac{dD}{di} = 1 + \frac{di'}{di} = 0. \tag{9}$$

On the other hand, differentiating the equation $e + e' = A$ with respect to e, we find

$$\frac{dA}{de} = 1 + \frac{de'}{de}. \tag{10}$$

From (8) and (9) we obtain $\frac{di'}{di} = \frac{de'}{de}$, whence

$$i = i' \tag{11}$$

and

$$e = e' = \frac{A}{2} \tag{11'}$$

as conditions for $D_{\min}$.

Further, from (9) and (10) we find that

$$D_{\min} = 2i - A \quad \text{or} \quad i = \frac{D_{\min} + A}{2}, \tag{12}$$

while from (10) and (11')

$$\sin \frac{D_{\min} + A}{2} = n \sin \frac{A}{2}.$$

We present the values of D_{min} as calculated from this formula for rays of different wavelengths (Table 85).

TABLE 85

Color	n	$A = 60°$	$A = 90°$
Violet	1.317	22°22'	47°16'
Yellow	1.310	21 50	45 44
Red	1.307	21 34	45 06

In the general case in which a ray incident at a certain angle h to the principal section of the prism is considered formula (12) assumes a somewhat more complicated form. If k is the angle between the ray and the principal section within the prism, we have

$$\sin \frac{D_{min}+A}{2} = n \frac{\cos k}{\cos h} \sin \frac{A}{2}, \qquad (12')$$

i.e., the values of the angle of minimum deflection change.

As one can see from (8), the angle of deflection D is different for different angles of incidence i. When i decreases i' increases and vice versa (of. Figure 178). Hence it follows that the ray experiences maximum deviation when the difference $i—i'$ reaches its maximum; this happens for $i = 90°$ or $i' = 90°$. Consequently,

$$D_{max} = 90 + i - A, \qquad (13)$$

whence

$$i = D_{max} + A - 90.$$

By applying the above formulas to different forms and arrangements of crystals in the atmosphere, one can explain the various halo phenomena. As illustration we will consider the formation of mocksuns and 22°-rings.

Let us suppose that the ice crystals are so oriented in the atmosphere that their principal axes are perpendicular to the horizontal plane and that the refracting angle $A = 60°$. Assuming that the rays are moving in the plane of the horizon, equation (12) gives the angle of minimum deviation D_{min} for which the intensity is maximum; for the violet and red rays the values obtained are those cited above. Consequently, at this distance from the sun we would obtain a colored spot 22°22'-21°34' = 48' wide, red on the inside, if the sun were a single point; but since the sun's diameter is 32', the width of the spot is 1°20'.

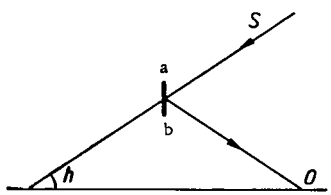


FIGURE 179

If the sun lies at a certain elevation above the horizon and therefore its rays fall at a definite angle to the plane of principal sections of the prisms, formula (12') must be employed. In this case we find that the distance of the 22°-parhelion from the sun increases with increasing solar elevation; at elevations greater than 60.8° parhelia cannot be observed at all.

Until now it was assumed that rays fell on the prisms at angles giving the minimum deviation, but in reality they fall at a wide variety of angles; for this reason bright tails extend from the mocksun and away from the sun, and this side has no sharp edge.

If one drops the proposition that crystals occupy oriented positions and assumes that their principal axes lie randomly, it is easy to understand that a bright ring colored red on the inside will be visible around the sun; its radius will be 21°6' for the red color and 22°4' for the violet. It will be sharply delineated on the inside and its intensity will fade gradually on the outside. This explains 22°-rings.

The explanation for halos and parhelia of 46° can be found in the same way, taking the refracting angle of the prisms as 90°.

The formation of colorless halo forms has a somewhat different explanation. Let us consider, for instance, the white horizontal circle, which is due to reflection of rays by the lateral faces of prisms.

Indeed, rays from the sun S falling on the face ab (Figure 179) are reflected by it, producing an image of the sun at the height O at which the sun occurs in the eye of the observer h (Figure 179). If the crystals lie in all possible directions from the observer, he will then see a white horizontal circle passing through the sun and extending over the entire sky.

All other halo forms, major ones as well as rare, unusual displays, can be explained in the same way by refraction, reflection, or a combination of the two.

§ 3. Coronas, glories and other analogous phenomena

If the sun and moon are observed through a thin veil of clouds, they will frequently appear framed by fairly bright iridescent rings called coronas. The inner portion of the corona — the so-called aureole — is contiguous with the sun (moon). It consists of a bright circle in which the colors grade from bluish-white through yellow to red, which lies on the outside. The dimensions of the aureole — of its red edges — vary considerably and sometimes reach 5°. Under favorable conditions the aureole is bordered on the outside by colored rings (up to three in number) separated by dark gaps. The colors in these rings, while less strong, occur in the same order as in the aureole (i.e., red is always on the outside). These phenomena are observed both in the presence of clouds with water droplet elements and in the presence of ice-crystal clouds.

The order of the colors and nature of the phenomenon indicate that we are dealing here with the diffraction of light, and the theory of coronas reduces to considering the diffraction of light moving through small holes or narrow slits. Water droplets act as holes and ice needles as slits.

Considering the case of the passage of monochromatic light through water drops and replacing the particle by a hole, we find that, as is known from optics, the distribution of intensity I in the diffraction pattern is expressed by the series

$$I = \pi^2 r^4 \left[1 - \frac{1}{2}\frac{m^2}{1} + \frac{1}{3}\frac{m^4}{(1\cdot 2)^2} - \frac{1}{4}\frac{m^6}{(1\cdot 2\cdot 3)^3} + \cdots\right], \qquad (14)$$

where

$$m = \frac{\pi r}{\lambda} \sin \theta$$

(θ is the angle of deflection of the rays from their initial direction and r is the drop radius).

The series (14) converges for all m and its sum for $m \geqslant 0$ gives the total distribution of intensity in the diffraction pattern. Calculations show that in the case of monochromatic light the pattern will consist of a series of rings separated by dark circles. The maxima of intensity and their relative values, and also the minima, which are equal to zero, are listed for different values of $\frac{m}{\pi}$ in Table 86.

TABLE 86

Order of ring	Maximum			Minimum		
	$\frac{m}{\pi}$	difference	I_{max}	$\frac{m}{\pi}$	difference	I_{min}
1	0.000		1	0.610		0
		−0.819			−0.506	
2	0.819		0.0174	1.116		0
		−0.527			−0.503	
3	1.346		0.0042	1.619		0
		−0.512			−0.502	
4	1.856		0.0016	2.121		0
		−0.504			−0.501	

The value of the angle θ_{min} corresponding to the intensity minima can be found from the relation

$$\sin \theta_{min} = \frac{\lambda}{r} \left(\frac{m}{\pi}\right)_{min}. \tag{15}$$

If one looks at the difference between values of $\frac{m}{\pi}$ corresponding to successive minima, one will see that it remains nearly constant and amounts to 0.50. This allows one to write the following empirical formula for the position of the minima:

$$\sin \theta_n = (n + 0.22) \frac{\lambda}{2r}, \tag{16}$$

where n is the order of the minimum.

From (16) and also from the table one can see that for small values of θ the first minimum ($n=1$) will have the angular radius

$$\theta_1 = 0.61 \frac{\lambda}{r} \tag{17}$$

and the first dark ring will lie here.

The second dark ring will have the angular radius $\theta_2 = 1.116 \frac{\lambda}{r}$. All subsequent rings will occur at intervals of

$$\theta_n - \theta_{n-1} = 0.50 \frac{\lambda}{r}. \tag{18}$$

Table 86 also shows that the intensity of the aureole (first maximum) is much greater than that of the subsidiary rings. It can be proved that in the case where not one but a multitude of identical holes is present, the

phenomenon retains the same character and only its intensity increases in proportion to the number of holes.

For coronas in clouds consisting of ice crystals (needles) the theory reduces to considering the diffraction of light through a narrow slit. Denoting the width of the slit by a, we obtain the following **expression** for the distribution of intensity:

$$I = a^2 \frac{\sin^2 u}{u^2}, \tag{19}$$

where

$$u = \frac{\pi a \sin \theta'}{\lambda}. \tag{20}$$

In this case the positions of the minima are determined by the conditions

$$\sin \theta_n' = n \frac{\lambda}{a}, \tag{21}$$

where $n = 1, 2, 3, \ldots$, and the positions of the maxima by the condition

$$u = \operatorname{tg} u. \tag{22}$$

Table 87 presents certain numerical data.

TABLE 87

Order of ring	Maximum			Minimum		
	$\frac{u}{\pi} = \frac{a \sin \theta'}{\lambda}$	difference	I_{max}	$\frac{u}{\pi} = \frac{a \sin \theta'}{\lambda}$	difference	I_{min}
1	0.000		1.000	1		0
		-1.430			-1	
2	1.430		0.0472	2		0
		-1.029			-1	
3	2.459		0.0165	3		0
		-1.012			-1	
4	3.471		0.0083	4		0
		-1.006			-1	

Comparison of Tables 86 and 87 shows that in the case of crystal clouds all the minima occur at equal intervals and that for the same intensity in the aureole the subsidiary coronas are brighter in the case of ice crystals.

Let us compare the size of the aureole in the case of water drops and ice needles, setting the width a of the latter equal to the drop diameter $(2r)$; in this case

$$\sin \theta' : \sin \theta = 1 : (1 + 0.22), \tag{23}$$

i.e., for water drops the aureole is 22% broader than it is for ice needles.

Since in reality light is not monochromatic and the source of light is not a point but has a certain extension; aureoles assume a rainbow coloring with blue-violet inside and red outside; owing to the superposition of the diffraction images of each point of the light source, the aureoles lose purity (becomes diffuse) and increase in width by the radius of the light source, i.e., by 16° for the sun and moon [sic].

Since the dimensions of coronas depend on the dimensions of the water drops and ice crystals, the former give a clue to the latter. As one easily sees from the formulas, the smaller the corona the larger the cloud elements involved. Prolonged observations lead to the conclusion that apparently many of the clouds giving rise to corona effects consist of ice

needles and that the elements of these clouds range mainly from 0.5 to $1.25 \cdot 10^{-3}$ mm in size.

Aside from corona phenomena, one may also observe a whole series of diffraction phenomena including: 1) Bishop rings, consisting of large coronas (up to 10° wide), red-brown in color and lying at distance of roughly 20 to 30° from the sun. It has been noted that Bishop rings are especially frequent after volcanic eruptions and also when the atmosphere contains large amounts of dust; 2) iridescent clouds, the edges of which assume a mother-of-pearl hue (usually observed in thin Ac and Ci clouds). The iridescent parts of the clouds lie at distances of several degrees to 20° and more from the sun.

A highly colorful phenomenon similar to coronas is the so-called glory. This phenomenon consists of a series of luminous colored rings (up to five) with red on the outside which, however, are seen against a cloud surface or fog wall on the antisolar (or antilunar) side with reference to the observer. The center of these rings lies along the line which is the continuation of the straight line connecting the sun and the observer. Glories are observed especially frequently in mountains at low solar elevations and during flights on aircraft and balloons, in which case they may form on cloud around the shadow cast by the aircraft or the head of the observer. In certain cases (in mountains) where glories form around the magnified shadow of the observer, this phenomenon is known as the Brocken Spectre (bow), after the Brocken mountains where it is particularly common. A phenomenon analogous to glories is observed around the shadow of the observer's head on the surface of meadows or fields, especially those covered with dew; it is known as aureole.

A satisfactory explanation of glories could not be obtained from the concepts of geometric optics; it was found only on the basis of Mie's theory. Without going into the highly complicated mathematical aspects of this theory, let us consider the essence of the phenomenon. In light scattering on very large water drops part of the light is scattered in the direction opposite to the incident ray; the scattering indicatrix presents a very complex appearance with a series of alternating maxima and minima. This diffracted light reaching the observer's eye from the cloud or fog particles is seen as a glory.

TABLE 88

Order of corona and glory	Maximum in corona, minimum in glory	Minimum in corona, maximum in glory
1	5.1	3.05
2	8.4	6.7
3	11.6	10.0

Mie's general theory of the scattering of light by a particle naturally includes the theory of coronas. It enables determination of the positions of the maxima and minima of light in the case of glories and in the case of coronas — though only for those composed of water drops and not ice particles. According to this theory, the angular separation of the maxima and

minima of coronas and glories is given by the expression

$$\sin\frac{\theta}{2} = b\frac{\lambda}{4\pi r}, \qquad (24)$$

where, as before, r is the drop radius and b a numerical factor; its values are given in Table 88.

It turns out that for the same drop size the positions occupied by maxima in coronas are occupied by minima in glories, and vice versa. Formula (24) is in better agreement with results of observations and experiments than (16). From this formula we have

$$r = \frac{b\lambda}{4\pi\sin\frac{\theta}{2}}. \qquad (25)$$

Thus the smaller the coronas or glories, the larger the diffracting drops responsible for their appearance.

Descriptions and detailed explanations of all the phenomena considered in the present chapter will be found in specialized textbooks of atmospheric optics.

Part Seven
ELECTRICAL PHENOMENA IN THE ATMOSPHERE (ATMOSPHERIC ELECTRICITY)

The electrical phenomena which take place in the atmosphere and the electrical properties of the atmosphere are important for many meteorological processes affecting our practical life.

The phenomena that are best known and susceptible of direct perception include electrical discharges in storms (lightning), luminous silent discharges from points (St. Elmo's fire) and auroras.

The wide range of problems associated with the study of the electrical properties of the atmosphere constitutes a separate branch of atmospheric physics— the study of atmospheric electricity. The principal results obtained in this field will be dealt with in the next few chapters.

Chapter 28

STATE OF IONIZATION OF THE ATMOSPHERE

§ 1. Fundamental concepts. Ions in the atmosphere and electrical conductivity

As early as 1755 it was discovered that a charged conductor placed in the air and perfectly insulated from the ground gradually loses its charge. The law according to which this loss of charge takes place states that the rate of depletion of charge is proportional at any instant to the charge Q on the conductor, i.e.,

$$\frac{dQ}{dt} = aQ, \tag{1}$$

which gives, after integration,

$$Q_t = Q_0 e^{-at}, \tag{2}$$

where t is the time and a the so-called decay coefficient.

The correct interpretation of this phenomenon was given only at the end of the nineteenth century when it was established experimentally that the atmosphere is in an ionized state, i.e., that it contains minute positively and negatively charged particles, which were named ions (radius $< 10^{-5}$ cm). Owing to the presence of ions the atmosphere is not an ideal insulator but rather possesses the ability to conduct electricity.

Atmospheric ions form when an energy sufficient to remove one of the outer valence electrons is imparted to a molecule or atom of atmospheric gas by an external agency. An electrically neutral atom in which the positive charge of the nucleus is equal to the total charge of the orbital electrons will become positively charged on losing one of the electrons. At normal pressure the electron released combines almost instantaneously (within less than 10^{-6} sec) with a neutral atom of the surrounding medium at the next higher level to form a negative ion. This gives rise to pairs of ions (positive and negative) having molecular dimensions and carrying a single elementary charge, equal to $e = 4.803 \cdot 10^{-10}$ CGSE $= 1.6 \cdot 10^{-19}$ coulomb.

However, these initially formed molecular ions are very short-lived (fractions of a second), as the influence of polarization forces results in

their combining with a certain number of molecules (about 10–15) from the surrounding air to form fairly stable complexes of molecules; these are known as normal, or light, ions. But the atmosphere always contains minute foreign particles of large size in a suspended state (condensation nuclei and other aerosol particles). The light ions combine with the latter, transferring to them their charge. This gives rise to larger ions, called heavy, or Langevin, ions. Medium-sized ions, the so-called medium or intermediate ions, can sometimes be detected in the atmosphere; their nature is not yet entirely clear.

As a rule heavy ions also carry a single elementary charge. A theoretical consideration of the question of their charge leads to the conclusion that they can have more than one elementary charge only if their radius is greater than 10^{-6} cm, but, as observations show, the number of such ions is insignificant.

The ions contained in the atmosphere can be treated as an admixture to atmospheric air differing from other air molecules and suspended particles only in the presence upon them of electric charges. Therefore in addition to all the forces commonly active in the atmosphere, the ions are also influenced by electrical forces. Under their influence ions occurring in an external electric field are displaced along the lines of force of this field at a velocity which is proportional to the field strength E and depends on the nature of the ion. The drift velocity in a field of unit strength is called the ionic mobility $k = \frac{u}{E}$, where u is the velocity of the ion. Usually mobility is expressed in cm/sec : V/cm, i. e., $cm^2/V \cdot sec$.

Mobility is the principal quantitative characteristic ions. It depends on the nature of the ionized gas. Thus the heavier and more complicated the molecules of the gas, the lower the mobility of the resulting ions. It also depends on the temperature and pressure of the gas, and varies in inverse proportion to its density.

The dependence of k on the temperature T and pressure p, which becomes very important in the case of ions in the upper atmospheric layers, can be represented as

$$k(T, p) = k(T_0 p_0) \frac{p_0}{p} \frac{T}{T_0}. \qquad (3)$$

From the kinetic theory of gases Langevin demonstrated that, in the first approximation, the mobility k of molecular ions is given by

$$k = a \frac{e}{m} \frac{l}{v}, \qquad (3')$$

where a is a certain numerical coefficient (of the order of 0.5–1.0), $\frac{e}{m}$ is the ratio of the charge of the ion to its mass, l is the mean free path of the ion and v is the mean velocity of its thermal motion.

Further development of Langevin's theory, by himself and by a number of other investigators, led to more exact but also far more complicated expressions for the mobility of light ions; the results of calculations from these expressions are in good agreement with observational data. According to laboratory studies, at room temperature (20°) and normal atmospheric pressure (760 mm) in pure air the mobility of light ions averages $k_+ = 1.37\ cm^2/V \cdot sec$ and $k_- = 1.89\ cm^2/V \cdot sec$; $\frac{k_-}{k_+} = 1.38$.

The mobility of light ions in the lower atmospheric layer under natural conditions also amounts to about $1-2\ cm^2/V\cdot sec$ but is closer to the value $1\ cm^2/V\cdot sec$; further, the mobility of negative ions is generally a little higher than that of positive ions, and is affected by atmospheric humidity; it decreases perceptibly when the humidity rises. With increasing elevation above sea level the mobility of light ions rises in conformity with (3).

The mobility of heavy ions is considerably (10^3-10^4 times) smaller than that of light ones and moreover varies widely.

The ions present in the atmosphere can be classified somewhat arbitrarily according to size and mobility, as in Table 89.

TABLE 89

Principal ion groups in the atmosphere

Ion group		Mobility k ($cm^2/V\cdot sec$)	Radius r (cm)
Light		>1	$6.6\cdot10^{-8}$
Intermediate	smaller	$1-10^{-2}$	$(6.6-80)10^{-8}$
	larger	$10^{-2}-10^{-3}$	$(80-250)10^{-8}$
Heavy (Langevin)		$10^{-3}-25\cdot10^{-4}$	$(250-550)10^{-8}$
Ultra-heavy		$<25\cdot10^{-4}$	$550\cdot10^{-8}$

We recall that the sizes of fog droplets and cloud elements are $10^{-4}-10^{-3}$ cm, while those of raindrops are even greater. These particles, like dust particles, can carry charges, but are not usually grouped among the ions.

The fundamental quantity which characterizes the state of ionization of the atmosphere is the number of ions per unit volume ($1\ cm^3$), known as the ion concentration. The number of ions in each of the two principal mobility groups, light and heavy, is usually considered separately. Sometimes the distribution of concentration over mobility is studied more closely and the distribution curve $n_k=f(k)$ is determined; this is the so-called spectrum of ionic mobilities.

The presence of ions in the atmosphere determines its conductivity. Indeed, ions move in an electric field of intensity E at the speed kE; as every ion carries a charge e, the charge $i_+=n_+k_+eE$ is transferred across one square centimeter of surface normal to the direction of the field E per unit time, and the charge $i_-=n_-k_-eE$ in the opposite direction. Their sum, given by

$$i=i_+ + i_- = (n_+ k_+ + n_- k_-)\, eE, \tag{4}$$

gives the density of the ionic current, while the products

$$\lambda_+ = n_+ k_+ e \quad \text{and} \quad \lambda_- = n_- k_- e \tag{5}$$

are the so-called polar conductivities; the sum of the polar conductivities gives the total conductivity

$$\lambda=\lambda_+ + \lambda_-. \tag{6}$$

Since the atmosphere contains ions with different mobilities, the expression for the conductivity should be written in the more correct form $\lambda = \sum_{j=1}^{\infty} (n_+ k_+ + n_- k_-)_j e$, where the summation extends over all mobility groups of ions in the atmosphere. If we confine ourselves to classifying the ions into three groups — light, intermediate and heavy, we may write

$$\lambda = n_l k_l e + n_m k_m e + N_h k_h e. \tag{7}$$

Using the mean values of the mobilities given in Table 89 we find that even for a large number of heavy ions N_h the first term in the above plays the leading part; calculations show that more than 95% of the conductivity of the atmosphere is due to light ions.

The decay of charge on a conductor is easily explained from the above considerations. Indeed, an electric field is set up around the charged body; its strength at the surface of the conductor s is related to the surface density of charge σ by $E = 4\pi\sigma$. The total charge of the body is $Q = \int_s \sigma ds = \int_s \frac{E}{4\pi} ds$. Under the influence of the field ions from the surrounding space create a current which for the entire surface of the conductor is

$$\frac{dQ}{dt} = -\int_s \lambda E ds. \tag{8}$$

From what we said it is evident that

$$-\frac{dQ}{dt} = \int_s \lambda E ds = 4\pi\lambda \int_s \frac{E}{4\pi} ds = 4\pi\lambda Q,$$

whence

$$Q = Q_0 e^{-4\pi\lambda t} = Q_0 e^{-at}, \tag{9}$$

i.e., the decay coefficient is related to the conductivity by

$$a_{\pm} = 4\pi\lambda_{\pm}. \tag{10}$$

Thus when a vertically directed electric field of strength E is present in the atmosphere, it will set up a vertical ionic current whose density, from (4), will be given by

$$i = i_+ + i_- = (\lambda_+ + \lambda_-) E = \lambda E. \tag{11}$$

This flux is usually directed towards the earth's surface. Since $\lambda = \frac{1}{r}$, where r is the resistivity of air, the above can be written as

$$i = \frac{E}{r}. \tag{11'}$$

If we consider a vertical column of air of unit cross section extending from the surface to the height h, the resistance of such a column is

$$R = \int_0^h r dh = \int_0^h \frac{1}{\lambda} dh.$$

Calculations show that the resistance R increases particularly rapidly with h in the lower atmospheric layers and already amounts to roughly $\frac{9}{10}$ of the total resistance of the entire atmosphere (10^{21} ohm/cm^2) in the 0 to 10 km layer.

In conclusion we note that due to the presence of ions and other charged particles in the atmosphere the amount of charge of any sign in a certain volume may vary. Thus a given volume may contain a certain net charge. The density of this space charge, i. e., the net charge per unit volume (cm^3 or m^3), is denoted by ρ. Space charges play a very great role in all phenomena of atmospheric electricity.

§ 2. Principal ionizing agencies in the atmosphere

The most important of the many known ionizing agencies in the lower atmospheric layers are the emanations of radioactive substances contained in the earth's crust and atmosphere, and cosmic rays. The ultraviolet radiation in the region of these wavelengths ($\lambda > 285\ m\mu$) which penetrate into the lower stratosphere and troposphere play no part whatsoever in the ionization of these layers. Their influence may be felt only in the photoelectric effect, but under natural conditions this effect is so small that it can be disregarded. In addition to the major ionizing agencies mentioned above, one may list many other factors; however, they are all of purely secondary importance and also very limited in space and time.

The ionizing effect of any agency in the atmosphere is measured by the number of ion pairs formed per second in a cubic centimeter of air at standard pressure and temperature. This unit is denoted by I.

The radioactivity of crustal rocks is very feeble. On the average rocks contain $1 \cdot 10^{-12}$ g of Ra per gram of rock, which corresponds roughly to $3 \cdot 10^{-6}$ g of U or 10^{-5} g of Th per gram of rock.

The radioactivity of ordinary rocks and surface waters is so low that the direct influence of their radiations is limited and extends to a small height only. The action of their α-emission, which penetrates into the atmosphere to a height of a few centimeters, can generally be disregarded altogether. Only the γ-radiation and, near the surface, the β-rays, play an appreciable role. The ionizing power of β-rays amounts to roughly $1I$ near the surface, and that of γ-rays to about $3I$. The ionizing effect of β-rays decreases exponentially with height and by a height of a few hundred meters can already be regarded as zero.

In view of the insignificant ionizing effect of these rays, it is the radioactive substances contained in the atmosphere itself that are chiefly responsible for ionizing the atmosphere. These substances are radioactive emanations (of radium, thorium, actinium) and their decay products.

Radioactive elements enter the atmosphere together with the soil air during its exchange with the outer air.

Investigations have established that the output of Ra emanations from the soil (exhalation) averages around $40 \cdot 10^{-18}$ curie/cm^2 sec *. On entering the atmosphere radioactive products are carried away from their source by vertical and horizontal currents over distances which increase as their lifetimes increase. Therefore from the standpoint of atmospheric activity the only substances of major importance are certain long-lived products (including the emanations of radium with a half life of 3.82 days), traces of which are still to be found in mid-ocean to heights of several kilometers

* We recall that a curie is the amount of radioactive substance which produces $3.7 \cdot 10^{+10}$ decays per second.

(up to 6 km). In Table 90 we present the basic data on crustal and atmospheric radioactivity.

TABLE 90

Radioactivity of earth's crust and atmosphere

Crustal rocks	$1-3\cdot10^{-12}$ g Ra/g
Soil air	$2\cdot10^{-11}$ curie/l
Atmospheric air:	
over land	$1.2\cdot10^{-13}$ curie/l
over sea at distances from shore	$1\cdot10^{-15}$ curie/l
over sea near shore	$1\cdot10^{-14}$ curie/l

The number of ion pairs formed in the atmosphere under the influence of radioactive emanations is nearly zero over the ocean far from the shore and amounts to several pairs (about 10/) overland near the surface, but varies from place to place. In areas of high radioactivity (mineral radioactive sources, birthplaces of radioactive elements, etc.) the radioactivity of the atmosphere and hence its ionization are naturally more vigorous.

In recent years radioactive substances formed during the testing of nuclear weapons have assumed very great importance. They are ejected at considerable altitudes and can remain suspended for long periods in the atmosphere, settling down to earth only very slowly. In view of their long half life certain of these (e.g., Sr^{90}, about 28 years) can give rise to supplementary ionization, not to speak of their harmful biological effects. The study of atmospheric contamination by artificial radioactive substances has lately received a good deal of attention and covers both the question of their concentration in the air and that of the amount settling on the surface (including the portion washed out of the atmosphere by precipitation).

The second ionizing agent is cosmic radiation, which is far more important for the ionization of air throughout the atmospheric layer. Cosmic rays, as the highly complex radiation reaching us from outer space is called, are chiefly of galactic origin, their sources being apparently the supernovas and possibly the novas.

It has now been established definitively that primary cosmic ray particles arriving from outer space consist chiefly of protons (99 %) and a small fraction of other particles having a tremendous energy (up to 10^{17} ev). When they penetrate into the atmosphere the primary particles interact with the atoms of the constituent gases and give rise to a great many different particles (electrons, positrons, photons, mesons, etc.), which in turn interact both with each other and with the atmospheric gases. As a result the flux of cosmic radiation accessible to observation at the earth's surface is highly complex and differs in composition from the flux of primary radiation entering the atmosphere. The intensity of the cosmic rays is nearly constant in time and the fluctuations that are recorded are completely negligible; only short-term variations of the nature of bursts are observed infrequently. These bursts coincide with the times of solar flares, which shows that the sun may, under specific conditions and at certain instances, act as a source of cosmic rays.

A very high penetrating power allows cosmic rays to travel through the entire thickness of the atmosphere and even to penetrate into the depths of the ocean and crust.

At the earth's surface the intensity of ion formation under the influence of cosmic rays is least near the equator, where it amounts to 1.6 I; as latitude increases it rises to 1.9 I at latitude 40°, then remains constant on further increase of latitude. Vertically the intensity of ionization due to cosmic rays increases with height, reaching a peak at about 12–18 km; the value at this maximum is greater, and its position somewhat higher, at greater magnetic latitudes. Figure 180 gives the curve of the intensity of ion formation due to cosmic rays at different heights at latitude 40°N., based on data of rocket observations. Recently, research with the aid of rockets and artificial satellites has yielded many new data on cosmic rays.

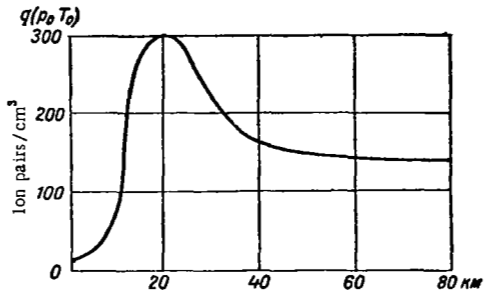


FIGURE 180. Intensity of ion formation under the influence of cosmic rays (latitude 41°N.)

If one compares the relative importance of the two ionizing agencies, one finds that overland at the surface 80% of the ionization is due to emanations of radioactive substances contained in the atmosphere, and some is due to γ-radiation of the earth's crust; in the layer up to 500 m 75% of the ionization is due to radioactive emanations; but the role of the latter decreases very rapidly and at heights above 5–6 km atmospheric ionization is already due almost entirely (100%) to cosmic rays alone. In total for the 0–90 km layer the latter give rise to 95% of the total number of ions formed.

Over the ocean at great distances from shore where atmospheric radioactivity is low the principal and for practical purposes the only ionizing agency is cosmic radiation. The same is true at great altitudes all the way up to the levels reached by the ultraviolet rays of the sun (wavelength $\lambda < 1000$ Å) and other forms of radiation (in particular corpuscular) possessing an energy great enough to ionize the gases. Starting from these heights in the ionosphere the principal ionizing agency is solar radiation.

In addition to the major ionizing agencies considered above, many other processes can lead to the formation of ions or charged particles in the atmosphere; however, they are of secondary importance and their significance is restricted in time and space. They include such factors as electrical discharges in thunderstorms, silent discharges from points, photoelectron emission from the surface of certain crystal minerals, thermal ionization in combustion processes and near active volcanos, balloelectric phenomena arising in the crushing of solid bodies and spraying of water, etc.

Frictional processes also play a certain part in the formation of charged particles in air; such processes are invariably observed when particles moving in air collide with one another or come into contact with a surface (in particular, the ground), particles with a greater dielectric constant taking on a positive charge and those with a smaller dielectric constant a negative charge. The charge assumed can be very considerable. Charged particles of smoke and dust lifted off the earth's surface by the wind, sand (during sandstorms) and ice and snow particles (during snowstorms) are formed in this way. But processes of formation of charges on cloud and precipitation particles are especially important and will be considered in detail later.

§ 3. Ion-destruction processes. Ion recombination

The concentration of ions of any given kind (mobility) and its variation in time are determined on the one hand by the number of ions formed per unit time and on the other by the number of ions annihilated in the same time. The most important of the ion-destruction processes is recombination; ions with opposite signs meet and neutralize one another.

Let us imagine, for illustration, that all ions contained in the atmosphere have the same mobility (e.g., light ions); we denote their concentration by n_+ and n_-.

Obviously, the number of ions recombining per second in one cubic centimeter of air is $\alpha n_+ n_-$, where α is a proportionality factor, the so-called recombination coefficient. If q pairs of ions are formed per second in one cubic centimeter, the change in the number of ions of each sign in time can be written as

$$\frac{dn_+}{dt} = \frac{dn_-}{dt} = q - \alpha n_+ n_-. \tag{12}$$

If we take $n_+ = n_- = n$ (this is fairly close to reality), we obtain

$$\frac{dn}{dt} = q - \alpha n^2. \tag{13}$$

Under conditions of stationarity, where $\frac{dn}{dt} = 0$, we have

$$q = \alpha n^2. \tag{14}$$

According to laboratory studies, in dry, completely pure air at normal pressure and temperature $\alpha = 1.6 \cdot 10^{-6}$ cm^3/sec for light ions. Observations under natural conditions for a pure atmosphere usually give a somewhat higher value ($\alpha = 1.6-4.6 \cdot 10^{-6}$ cm^3/sec).

Theoretical treatment of the question of recombination shows α to be dependent on the kind of gas involved as well as on pressure and temperature.

From (14) it follows that under conditions of stationarity the ion concentration is $n = n_\infty = \sqrt{\frac{q}{\alpha}}$. Integrating (13), we obtain the ion concentration n_t at time t subsequent to the cessation of ionization ($q=0$) in the state of equilibrium:

$$n_t = \frac{n_\infty}{1 + \alpha n_\infty t}.$$

In the same way one can find the ion concentration at time t subsequent to the onset of ionization:

$$n_t = \sqrt{\frac{q}{\alpha}}\left[\frac{1-e^{-2\sqrt{\alpha q}\,t}}{1+e^{-2\sqrt{\alpha q}\,t}}\right].$$

It should be noted, however, that the above relations can be applied (and that too only approximately) only to a particularly pure atmosphere, e. g., over oceans and at considerable altitudes. Indeed, the atmosphere always contains not only light ions but also ions with other mobilities, in particular heavy ions, as well as uncharged nuclei. The concentration of these large particles is considerably greater than that of light ions. One must therefore consider the recombination of all ions of opposite signs and various mobilities, for instance light with heavy and medium, heavy with each other, and so forth. In addition, one must allow for the settling (adhesion) of ions on larger neutral particles suspended in air, which results in the conversion of ions of any given mobility (mainly light) into heavier ones. All such processes of recombination of light and heavy ions and settling of light ions on uncharged nuclei rise to the forefront when one considers the question of the variation of the light-ion concentration in the lower, dust-filled layers of troposphere. They can be allowed for approximately by introducing a further term $\beta' nN$ into the right-hand side of equation (13), where β' is a certain coefficient having the significance of a recombination coefficient, and N is the total number of larger particles, charged as well as uncharged. Then for the given case we write

$$\frac{dn}{dt} = q - \alpha n^2 - \beta' nN. \tag{15}$$

Indeed, let us confine our attention to the two processes just mentioned (recombination and adhesion) and consider the variation in time of the number of light ions n_+, n_- and heavy ions N_+, N_-, and of the number of uncharged particles N_0. We can then write down the following five equations:

$$\left.\begin{aligned}
\frac{dn_+}{dt} &= q - \alpha n_+ n_- - \eta_{+,-} n_+ N_- - \eta_{+,0} n_+ N_0 \\
\frac{dn_-}{dt} &= q - \alpha n_- n_+ - \eta_{-,+} n_- N_+ - \eta_{-,0} n - N_0 \\
\frac{dN_+}{dt} &= Q_+ + \eta_{+,0} n_+ N_0 - \eta_{-,+} n_- N_+ - \gamma N_- N_+ \\
\frac{dN_-}{dt} &= Q_- + \eta_{-,0} n_- N_0 - \eta_{+,-} n_+ N_- - \gamma N_+ N_- \\
\frac{dN_0}{dt} &= Q_0 + \eta_{-,+} n_- N_+ + \eta_{+,-} n_+ N_- + \\
&\quad + 2\gamma N_+ N_- - \eta_{+,0} n_+ N_0 - \eta_{-,0} n_- N_0
\end{aligned}\right\} \tag{16}$$

The meaning of these equations is obvious. Here q_+, q_-, Q_+, Q_- and Q_0 denote the number of ions and neutral particles produced per unit volume due to the action of all processes, e. g., ionization, diffusion, advection, etc.; α is the recombination coefficient for light ions and γ the same for heavy ions; we note that the numerical value of γ is three orders smaller than that of α ($\gamma \approx 1.4 \cdot 10^{-9}\,\mathrm{cm^3/sec}$); η is the coefficient of recombination between light and heavy ions, while $\eta_{+,0}$ and $\eta_{-,0}$ are the coefficients of adhesion of light ions to neutral nuclei.

The values of all the η-coefficients, as the theory (after Bricard) shows, depend on the diffusion coefficients of light ions of each sign and also on the effective radius (size) of the neutral particle and correspondingly of the heavy ion, increasing as the latter increases. Experimental data show that their values are of the same order of magnitude as the recombination coefficient of light ions ($\approx 10^{-6}$ cm^3/sec); they are somewhat different at different places and vary with time. According to observations at Pavlovsk (Leningrad area) their mean values are: $\eta_{+,0} = 4.4 \cdot 10^{-6}$, $\eta_{-,0} = 4.1 \cdot 10^{-6}$, $\eta_{+,-} = 6.5 \cdot 10^{-6}$ and $\eta_{-,+} = 6.4 \cdot 10^{-6}$ cm^3/sec.

However, even in approximate form the utilization of equations (16) for describing the state of ionization of the atmosphere is an exceedingly complicated matter, since it calls for the determination of a great many quantities which are difficult to measure and moreover highly variable in time. Usually, therefore, the system (16) is simplified still further and is used to study the state of equilibrium which for constant ion formation and annihilation usually sets in within a few minutes (up to 15 min).

If one disregards the recombination of heavy ions among themselves (since $\gamma \ll \alpha$), sets $Q_+ = Q_- = Q_0 = 0$, takes $q_+ = q_- = q$ equal to the number of light ions produced solely as a result of ionization, and further assumes that $n_+ = n_- = n$, $N_+ = N_- = N'$, $\eta_{+,-} = \eta_{-,+} = \mu$ and $\eta_{+,0} = \eta_{-,0} = \eta$, the first two of equations (16) simplify to

$$\frac{dn}{dt} = q - \alpha n^2 - \mu n N' - \eta n N_0 = q - n\,[\alpha n + \mu N' + \eta N_0], \qquad (17)$$

or, in brief,

$$\frac{dn}{dt} = q - \beta n, \qquad (18)$$

where

$$\beta = \alpha n + \mu N' + \eta N_0. \qquad (19)$$

Using (16) with the above simplifications, one can show that

$$\beta = \alpha n + \beta' N, \qquad (19')$$

where $N = N_+ + N_- + N_0$ is the total number of charged and uncharged nuclei per unit volume and $\beta' = \frac{2\mu\eta}{2\eta + \mu}$. The quantity β' is of the same order as α ($\sim 10^{-6}$ cm^2/sec).

As a result we arrive at the general (approximate) equation (15), which describes the change in the number of light ions in time. This equation reduces to (13) for a pure atmosphere where $N \approx 0$ and to (18) for a very dusty one where $N \gg n$.

The quantity β is called the light-ion dissipation constant (after von Schweidler) and characterizes the disappearance of light ions owing to adherence to neutral aerosol particles and recombination. Since α and β' are numerically very close, if $N \gg n$, as is the case for dust-filled air, one can assume approximately that $\beta = \beta' N$; clearly, the values of β depend heavily on the total number of nuclei N. Observations carried out at a number of points give average values of β of the order of 10^{-3} cm^3/sec — from a few units to several tens (5–100) of 10^{-3}.

Integrating (18), we obtain the number of ions n_t at the time t subsequent to the cessation of ionization ($q = 0$); it is $n_t = n_\infty e^{-\beta t}$. From the expression obtained one can see that the constant β can be regarded as analogous (only formally, it is true) to the constant of radioactive decay. This

provided some ground for introducing a "mean lifetime" of the light ion, defined as the average time interval $\tau=\frac{n}{q}$ between its formation and its disappearance. If only light ions are present then $\tau=\frac{n}{\alpha n^2}=\frac{1}{\alpha n}$; in the presence of a large number of nuclei and heavy ions $\tau=\tau'=\frac{1}{\beta}$. Observations and calculations show that the mean lifetime of light ions in the atmosphere varies according to the purity of the air between roughly 30 sec (for dust-filled air) and 1000 sec (for pure air). As to the lifetime of heavy ions, it is many times longer than that of light ions and may be as much as an hour or more.

When considering the ion content of the atmosphere one must always bear in mind that at any given point the concentration changes not only by virtue of the processes considered above, but also due to transfer under the influence of electrical forces, diffusion from regions of higher concentration and mechanical transport by moving air masses. This explains the great variability of the ion content and the very considerable difficulties in a theoretical treatment of this question. Direct measurements are thus of great value.

§ 4. Ion concentration and conductivity of the atmosphere according to observational data

The conductivity of the atmosphere and the ion concentration are most commonly measured by sucking air through a cylindrical condenser. If an electric field is set up inside such a condenser, for example by raising the internal electrode CD to a potential V, then an ion K (Figure 181)

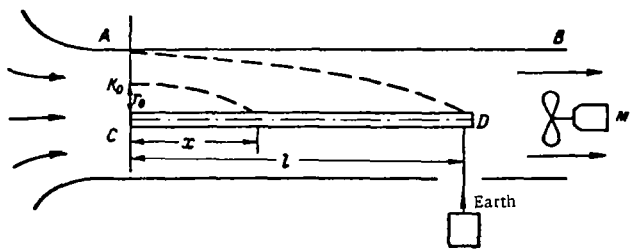


FIGURE 181. Instrument for measuring ion number and conductivity

entering the condenser at the distance r_0 from its axis will travel inside the latter under the influence of the mechanical force of air suction and the electrical force directed at right angles to the axis of the condenser; its trajectory will be a parabola. Depending on the rate of suction M (cm^3/sec) of air, the capacity C (cm) of the condenser and the potential difference between its plates, all ions with mobility greater than a certain limiting value

$$k_{\lim}=\frac{M}{4\pi CV}. \qquad (20)$$

and a certain proportion of ions with lower mobility will be captured.

By measuring the ion current set up inside the condenser one can determine both the conductivity of the air and the ion concentration. The appropriate methods are described in special handbooks; in particular the reader is referred to a monograph by I. M. Imyanitov.

On the basis of numerous observations at different points, one can take the average concentration of light ions near the surface to be approximately 400–500 pairs per cubic centimeter. The number of positive ions n_+ is a little larger in most cases than the number of negative ions n_-, roughly in the ratio $\frac{n_+}{n_-} = 1.10–1.20$. Correspondingly the conductivity of the atmosphere, which is determined, as we saw, chiefly by the light ions, amounts to about $2.2 \cdot 10^{-16}$ ohm^{-1} cm^{-1} $= 2 \cdot 10^{-4}$ CGSE, the ratio $\frac{\lambda_+}{\lambda_-}$ being likewise greater than unity. Table 91 gives certain summarized (tentative) data. At individual points the values of n and λ may deviate considerably from the mean values given in the table; the range of variation is indicated in parentheses.

TABLE 91

Mean conductivity and number of light ions near the earth's surface (after Mühleisen)

Area	Number of light ions (n/cm^3)	Conductivity λ (ohm^{-1} cm^{-1})	Daily fluctuations
Large town	150 (30–1000)	$0.5 \cdot 10^{-16}$	Large
Small town	250 (50–1000)	$1 \cdot 10^{-16}$	Medium
Rural	500 (100–1000)	$2 \cdot 10^{-16}$	Small
Ocean	600 (300–1000)	$2.4 \cdot 10^{-16}$	Almost none

As to the concentration of the heavy ions N_+ and N_-, this varies within such a wide range that it is difficult to pinpoint any mean value. One may state that the overland concentration of heavy ions at the surface is considerably higher (by a factor of 10–100) than the concentration of light ions, and that it varies over an even broader range (few hundreds to several tens of thousand) from place to place; its variation in time is also stronger.

In most places the $\frac{N_+}{N_-}$ ratio, like the corresponding ratio for light ions, is greater than unity, and close to 1.10 on the average. The concentration of light ions, heavy ions and neutral particles and the conductivity are closely linked; this is expressed in the fact that when the number of particles suspended in the atmosphere increases, the number of light ions decreases and the number of heavy ions increases.

In addition, one may assume that of all the particles N contained in the atmosphere on the average roughly half are charged and form heavy ions. The ratio of the number of uncharged particles to the number of particles of any one sign is on the average close to 2, increasing somewhat as the total number of particles increases and vice versa.

As to the intermediate ions, their concentration is highly variable.

Observations of the distribution of ionic mobilities (ion spectrum) reveal extreme variability and a high degree of dependence on local conditions and the state of the atmosphere.

For light ions a daily periodicity with main peak in the late night and early morning hours and minimum in the pre-noon hours is recorded at most observation points.

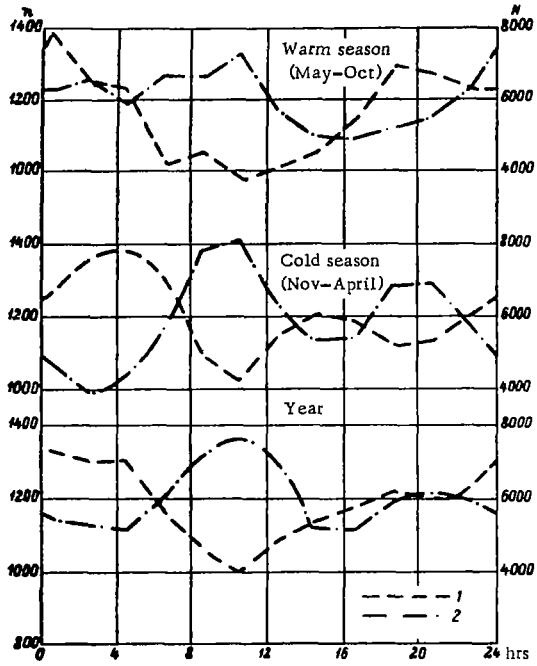


FIGURE 182. Daily march of concentration of light (1) and heavy (2) ions at Pavlovsk

An example is the daily march of the ion number at Pavlovsk (Figure 182). The morning peak is due to the maximum air purity at these hours. A secondary maximum in the afternoon hours and secondary minimum in the evening are often distinguished at a number of places, including Pavlovsk. Over oceans daily variations are very limited. For heavy ions the daily pattern is reversed.

The annual march of the light-ion concentration is fairly complicated and varies considerably from place to place. At many points the maxima are observed in the warm half of the year and the minima in winter. In general, however, the annual march is not particularly well developed and is largely dependent on local conditions and the state of the atmosphere.

The conductivity of the atmosphere $\lambda_{\pm} = n_{\pm} k_{\pm} e$ depends not so much on the ion number as, to a far greater extent, on their mobility. The periodicity of atmospheric conductivity is therefore roughly similar to that of the light-ion number but not quite identical with it. In the annual march, which depends strongly on local conditions, the atmospheric conductivity

at most middle-latitude points is greater in the warm half of the year than in the cold half.

As to the daily march of the conductivity, a peak in early morning (stronger in summer) and minimum in the evening is observed at most middle-latitude points of the northern hemisphere (Figure 183). This typical pattern is observed, for instance, at Pavlovsk. In some areas a secondary maximum in the afternoon hours, as at Tashkent, for instance, is also recorded. The same pattern as at Tashkent is observed on the island of Java. Over oceans the conductivity changes very little during the day.

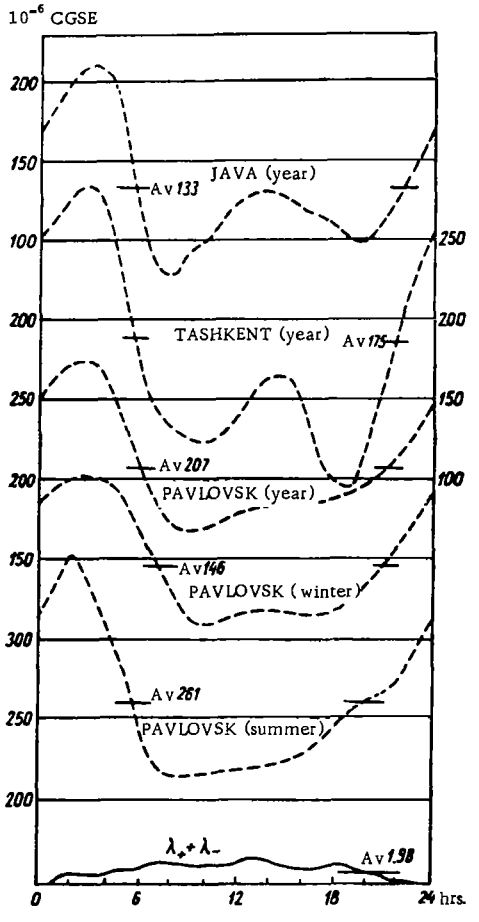


FIGURE 183. Daily march of conductivity at various points and over Pacific (lower curve)

The morning peak is the principal characteristic feature of the daily march of conductivity and can be attributed to the fact that at this time of day the atmosphere is least contaminated and the number of light ions and their mobility are therefore greater.

The conductivity and also the concentration and mobility of ions depend to a considerable extent on weather conditions. When the weather changes

they experience sharp and irregular fluctuations. The link between the state of ionization of the atmosphere and the degree of dustiness of the air is particularly close; basically it accounts for the links with other meteorological phenomena. Low conductivities, small light-ion concentrations and large heavy-ion concentrations are observed during haze and fog. For the same reason there is a close correlation with the visibility, deterioration of which reduces the concentration of light ions and the conductivity while increasing the number of heavy ions.

All these relationships can frequently be identified with ease in each individual case by analyzing the fundamental relation (16) which relates the intensity of ion production to the factors which govern the disappearance of light ions. These factors also control the character of the variation of the state of ionization of the atmosphere with height.

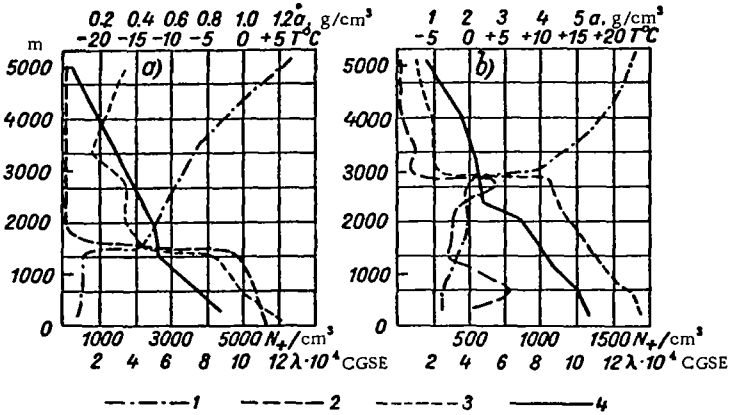


FIGURE 184. Vertical distribution of conductivity λ (1), heavy-ion concentration N (2), absolute humidity a (3) and temperature T (4) during flights over New England, 15 June 1953 (a) and 18 August 1953 (b)

Early conductivity measurements from high mountains, balloons and aircraft already revealed a general increase in conductivity with height in the troposphere. Recent observations have made it possible to obtain data to a height of about 30 km. These confirm that conductivity increases with height but indicate that this increase becomes fairly regular only starting from a certain height (roughly 3 km). Below 3–4 km the conductivity varies very irregularly owing to the presence of layers of dust and cloud in which the conductivity drops sharply; the vertical rise in conductivity is therefore weakly expressed in the lower layers, another factor contributing to irregularity being the more vigorous mixing prevalent in these layers.

Results of measurements of the conductivity and heavy-ion concentration over New England are presented for illustration in Figure 184. Above the layer of irregular variation a fairly smooth rise sets in, interrupted at rare intervals by cloud layers, and at a height of about 10 km the conductivity has increased roughly 10–12 times compared with its ground values. Above 10 km the conductivity continues to rise with height, as one can see from

Figure 185, which presents data obtained in New Mexico (solid line) with the aid of radiosondes. For the sake of comparison the figure also shows

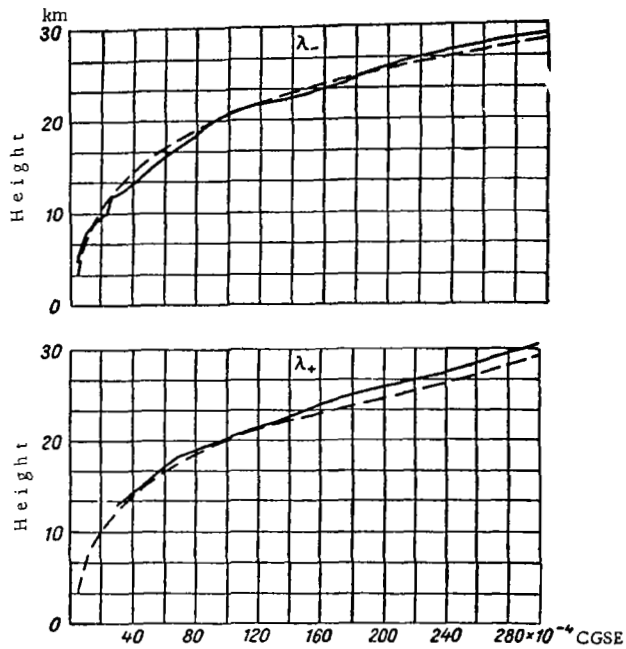


FIGURE 185. Conductivity, after observations in New Mexico

the vertical variation of conductivity as calculated under the assumption that cosmic rays are the only ionizing agency at these altitudes (dashed curves). Such calculations can be made from the formula

$$\lambda = \lambda_+ + \lambda_- = (n_+ k_+ + n_- k_-) e \approx 2nke,$$

where n and k are respectively the number of ions and their mobility at the height under consideration.

The number of ions of each sign at a certain height $n(h)$ can be calculated from a formula which follows from (14) and has the form

$$n(h) = \sqrt{\frac{q(h)}{\alpha(h)}},$$

where $q(h) = q_0(0) \frac{p}{p_0} \frac{T_0}{T}$ is the intensity of ion formation, which must be known, and $\alpha(h)$ is the recombination coefficient.

TABLE 92

Vertical variation of conductivity

h (km)	0	1	3	6	9	12	15	20	30
λ (CGSE)	$2 \cdot 10^{-4}$	$6 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	$1 \cdot 10^{-2}$	$2 \cdot 10^{-2}$	$8 \cdot 10^{-2}$	1

The mobility of ions at a given height $k(h)$ can also be determined from formula (3).

Figure 185 shows that results of calculations agree fairly closely with data of direct measurement.

Table 92 gives tentative calculated data on the variation of the conductivity λ with height.

Chapter 29

THE IONOSPHERE

§ 1. Ionosphere. Formation of the ionized layer

The rise in atmospheric conductivity with height, established by direct measurements to about 30 km, continues even at greater altitudes. This is because, aside from the cosmic rays, new and far more powerful ionizing agencies come into play at these heights; they include first and foremost the ultraviolet ($\lambda < 0.1\,\mu$) radiation of the sun, its corpuscular radiation, and meteors and stellar radiation. Further, at heights at which the density of the atmosphere is very low electrons formed in the process of ionization can exist for long periods in the free state, producing a high atmospheric conductivity. This ionized region of the atmosphere is known as the ionosphere.

The ionosphere begins at roughly 60–70 km, at which height the electron concentration N_{el} reaches values of about $(1-3)\cdot 10^3$ free electrons per cubic centimeter (el/cm^3) in daytime.

A particularly rapid rise in conductivity occurs at a height of about 100 km, where it is roughly 10^{12} times greater than at the surface. Higher up the electron concentration continues to increase, varying in a more or less complex manner, and reaches maximum values at a height which varies between 250 and 450 km depending on the time of day, latitude and longitude. At this principal maximum N_{el} amounts to about $2\cdot 10^6$ el/cm^3, occasionally reaching a peak of about $5\cdot 10^6$ el/cm^3. As one rises further there is a slower decrease in N_{el} until very great altitudes are reached, and at a height of the order of 2000 km ionization still amounts to 10^3 el/cm^3, gradually passing over to the concentration characteristic of an ionized interplanetary medium. A schematic diagram showing the variation of N_{el} with height above the earth's surface is given in Figure 186. Although systematic study of the ionosphere has been in progress for over thirty years, our information regarding it is altogether too limited for the construction of a complete theory of its formation, and many problems relating to its structure still remain unresolved.

The formation of a layer of high ionization in the upper atmosphere can be understood on the basis of the following elementary considerations. Indeed, if the atmosphere consisted of any one single gas then, since density in the atmosphere decreases with altitude while the intensity of the sun's ionizing radiation decreases as it penetrates into the atmosphere, at a certain height the degree of ionization would obviously reach maximum values, decreasing both upwards and downwards from this maximum. These considerations furnished the basis for the development of the so-called simple layer theory which, while though far from giving a complete

picture of the ionosphere, indicates the roles of the major factors involved in its formation.

Let us imagine that the atmosphere consists of a single arbitrary gas and that its density decreases with the height z according as

$$\rho = \rho_0 e^{-\frac{z}{H}} = \rho_0 e^{-\frac{gz}{RT}}, \tag{1}$$

where $H = \frac{RT}{g} \frac{R^*T}{\mu g}$ is the height of the homogeneous atmosphere; R is the gas constant for the given gas, T is the mean atmospheric temperature and g is the acceleration of gravity.

Consider a plane layer of the atmosphere ionized by solar radiation of intensity S incident at the zenith angle ζ* (Figure 187). On the element of

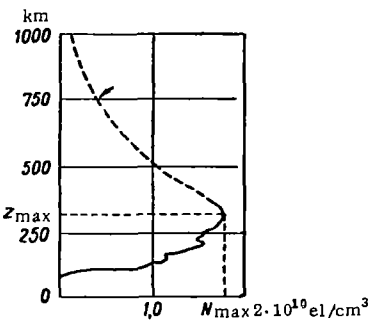


FIGURE 186. Vertical variation of electron concentration in the ionosphere

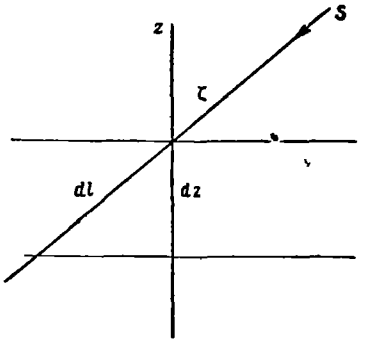


FIGURE 187

path $dl = \frac{dz}{\cos\zeta}$ the reduction in the intensity of the radiation will obviously be

$$dS = AS\rho\, dl = AS\rho_0 e^{-\frac{z}{H}} \frac{dz}{\cos\zeta}, \tag{2}$$

where A is the molecular absorption coefficient for the given radiation per unit mass of gas.

If we denote the initial intensity of the radiation before its entry into the absorbing layer by S_∞, then, integrating the last equation, we obtain

$$S = S_\infty e^{-\frac{A\rho_0 H}{\cos\zeta} e^{-\frac{z}{H}}} \tag{3}$$

If β is the number of ions formed as a result of the absorption of one unit of radiation, the rate of ion formation $q(z, \zeta)$ will be

$$q(z, \zeta) = \beta \frac{dS}{dz} \cos\zeta, \tag{4}$$

or, with (2) and (3),

$$q(z, \zeta) = \beta A S_\infty \rho_0 e^{-\frac{z}{H} - \frac{AH\rho_0}{\cos\zeta} e^{-\frac{z}{H}}} \tag{5}$$

* $\zeta < 85°$, if the curvature of the atmosphere be taken into account.

The maximum number of ions q_{max} $(z\zeta)$ formed per cubic centimeter per second at height z and zenith distance ζ can be found from the condition

$$\frac{dq(z, \zeta)}{dz} = 0,$$

which, from (5), can be written as

$$\frac{z_{max}}{H} = \ln \frac{A\rho_0 H}{\cos \zeta},$$

or

$$z_{max} = H \ln \frac{A\rho_0 H}{\cos \zeta}. \quad (6)$$

The maximum intensity of ion formation at this height will be

$$q_{max}(\zeta) = \frac{\beta S_\infty \cos \zeta}{He}. \quad (7)$$

Denoting $q_{max}(\zeta)$ and z_{max} for normal incidence $(\zeta=0)$ by q_0 and z_0 respectively, we obtain from (6) and (7)

$$z_0 = H \ln A\rho_0 H \text{ and } q_0 = \frac{\beta S_\infty}{He}. \quad (8)$$

Consequently,

$$q_{max}(\zeta) = q_0 \cos \zeta. \quad (8')$$

Then as a result of simple transformations expression (5) takes the form

$$q(z, \zeta) = q_0 \exp\left(1 + \frac{z_0 - z}{H} - \frac{1}{\cos \zeta} e^{-\frac{z_0 - z}{H}}\right), \quad (9)$$

which gives the height distribution of the intensity of ion formation for given ζ.

Introducing a certain effective electron recombination coefficient α, the change in the number of the electrons in time can be written as

$$\frac{dN_{el}}{dt} = q - \alpha N_{el}^2. \quad (10)$$

Under equilibrium or quasi-equilibrium conditions, in which $\frac{dN_{el}}{dt} \approx 0$, we have

$$N_{el} = \sqrt{\frac{q}{\alpha}}, \quad (11)$$

which in view of (9) gives

$$N_{el} = N_{max} \exp\left(\frac{1}{2}\left[1 + \frac{z_0 - z}{H} - \frac{1}{\cos \zeta} e^{-\frac{z}{H}}\right]\right), \quad (12)$$

where, as is easy to see from (8'),

$$N_{max} = N_0 \sqrt{\cos \zeta}. \quad (13)$$

The above shows that the electron concentration at the height of maximum ionization should be proportional to $\sqrt{\cos \zeta}$, and therefore in the daylight hours the peak of N_{el} occurs at noon and over the year at the summer solstice.

From the above it is evident that the vertical distribution of electron concentration (12), height of occurrence of the maximum (6) and value of

the electron concentration at this maximum (3) depend on the nature of the gas, the distribution of temperature in the atmosphere, the intensity and energy of the ionizing radiation S_∞, the electron recombination coefficient α and the zenith distance of the sun.

In the real atmosphere, which is a mixture of different gases and in which several ionizing agencies may act simultaneously, the structure of the ionosphere is of course far more complex than would seem to follow from the scheme just considered.

Early experimental data on the ionosphere obtained by observing the refraction and reflection of radio waves have now been supplemented by new data thanks to the launching of artificial earth satellites and geophysical rockets. Revision of the state and structure of the atmosphere at these altitudes will doubtless lead to a more complete theory of the ionosphere

§ 2. Propagation of electromagnetic waves in the ionosphere

It is well known that when electromagnetic waves are propagated in a medium which contains electrons the latter are excited into oscillatory motion under the influence of the electric field of the propagated wave. The frequency of oscillation is the same for all electrons as the frequency of the electric field. The phases are different for different electrons and are determined by their distance from the primary source. As a result both the refractive index μ and the damping constant δ change in the ionized medium; these quantities depend significantly on the electron concentration N_{el}, the frequency of the propagated wave ω and the frequency of collision ν between moving electrons and surrounding particles. As a result of the collisions the energy of oscillatory motion of the electrons transforms into thermal energy and the energy of the electromagnetic waves propagating through the medium becomes absorbed. As a consequence of this, if the refractive index of electromagnetic waves in vacuum is unity, in an ionized medium it will be less than unity and, as the theory shows, will be given by

$$\mu = \sqrt{1 - \frac{4\pi N e^2}{m} \frac{1}{\omega^2 + \nu^2}}, \tag{14}$$

while the absorption coefficient

$$\delta = \frac{2\pi N \frac{e^2}{m} \frac{\nu}{\omega^2 + \nu^2}}{c_0 \sqrt{1 - \frac{4\pi N e^2}{m} \frac{1}{\omega^2 + \nu^2}}}, \tag{15}$$

where e and m are the charge and mass of the electron, c_0 is the speed of light and the remaining symbols are as above.

Thus one can state that for electromagnetic waves the ionosphere is a medium with a smaller optical density than the underlying nonionized layers of the atmosphere. Therefore an electromagnetic ray moving into the ionosphere will be refracted, deviating from the normal to the surface of discontinuity between the layers; this deviation, moreover, increases as the electron concentration increases. If the electron concentration increases continuously with height in the ionized layer, the electromagnetic rays entering it will propagate inside it along a curve concave towards the

earth's surface, and if at a certain height the electron concentration proves large enough the rays will be deflected from the normal by 90° and, being reflected, will plunge back to the earth's surface. If the electromagnetic wave is sent upwards vertically (as is usually the case in practice), reflection will take place at sharp changes in the refractive index with height and will occur at the point where the refractive index becomes zero, i.e., where

$$\mu = \sqrt{1 - \frac{4\pi Ne^2}{m}\frac{1}{\omega^2 + \nu^2}} = 0. \tag{16}$$

If the electron collision frequency ν in the ionospheric layers under consideration is much smaller than the frequency of the propagating electromagnetic waves ($\omega^2 \gg \nu^2$), the above expression can be written as

$$1 - \frac{4\pi Ne^2}{m\omega^2} = 0, \tag{16'}$$

whence, introducing the frequency in hertz $f = \frac{\omega}{2\pi}$, we obtain

$$N = \frac{m\omega^2}{4\pi e^2} = \frac{m\pi}{e^2} f^2, \tag{17}$$

or, numerically for electrons,

$$N = 1.24 \cdot 10^{-8} f^2 \text{ el/cm}^3. \tag{17'}$$

The frequency f which determines, according to the above, the electron concentration for which waves of this frequency will be reflected is termed the critical frequency f_{cr}. This relation underlies the pulse method of mapping the high-frequency characteristics which give the distribution of the critical frequencies with height. The pulse method is the one most frequently employed to study the ionosphere. Characteristics can be obtained by sending out consecutively short pulses of evenly increasing frequency (usually within 0.5–15 Mhz) into the ionosphere and receiving the reflected signals. The values of $(N_{el})_{max}$ are found from the values of f_{cr} for which reflection ceases, and the level of the maximum electron concentration from the time taken by the signal to travel from the earth to the point of reflection and back. By comparing the amplitudes of the direct and the reflected pulses, one can also find the absorption coefficient δ.

Obviously, this method does not allow investigation of layers situated above the maximum concentration, since reflection is not observed from these owing to the decrease in electron concentration. They can be studied only by observing the propagation of radio waves emitted from a source located above the layers under investigation. Such possibilities have arisen in the study of the propagation of radio waves emitted from rockets or satellites.

A number of general conclusions concerning the propagation of electromagnetic waves through the ionosphere as determined by their absorption and reflection (refraction) can be drawn from the above discussion. Only waves with a small wavelength for which the refractive index is only slightly different from unity can move through the ionosphere without pronounced deviation of their path. Calculations show that, for instance, light waves ($\lambda \approx 0.5\mu$) cross the ionosphere freely, except for those (in the region $\lambda < 0.1\mu$) which are absorbed and induce ionization without reaching the surface. As the wavelength increases (in the radiofrequency region) the

deflection of the path becomes appreciable: for maximum values of the electron concentration observed in the ionosphere (about $5 \cdot 10^6$ el/cm^3), the paths of 10-m waves, for which the refractive index is roughly 0.77, are already strongly deflected, and the waves undergo reflection. Waves longer than 20 m do not propagate at all in the ionosphere since their refractive index is imaginary.

At the same time, the propagation of electromagnetic waves is strongly influenced by absorption. Since the absorption coefficient (cf. formula (15)) increases in proportion to the product $N\nu$ and decreases as the particle grows, in the denser layers where ν is high there will be considerable absorption of radio waves, especially those of low frequency. Calculations show that for the maximum value of $N\nu$ possible for the ionosphere (about 10^{10}) the damping of electromagnetic waves becomes especially pronounced only starting from wavelengths of tens of meters and over; for example, for 10-m waves damping by 10% would take place on a path of 300 km, whereas for light waves it would require a path length of 10^{15} km.

Thus longer waves (wavelength tens of meters) cannot penetrate into the ionosphere since they are reflected already in the lowest layers of the ionosphere. Short waves penetrating into the ionosphere will be deflected; ten-meter waves ($\omega = 1.8 \cdot 10^8$ sec^{-1}) will experience marked bending only in those layers of the ionosphere where $N_{el} > 10^5$ el/cm^3. Only ultrashort waves ($\lambda < 10$ m) will be able, for definite values of N_{el}, to penetrate through the atmosphere and reach beyond its boundaries.

§ 3. Structure of the ionosphere according to observational data

It has been established by numerous observations, chiefly with the aid of special ionosphere stations, that the general backdrop illustrated schematically in Figure 186, in which the electron concentration rises with height, is spotted by fairly stable formations of high concentration on which reflection of radio waves sent out from the earth's surface is regularly observed. These regions were named ionospheric layers; the most important of these in order of increasing altitude are the D-, E- and F- (F_1 and F_2) layers.

The D-layer lies at a height of 60–70 km. The maximum electron concentration inside it is estimated at $1-3 \cdot 10^3$ el/cm^3. At the same time the number of electron collisions ν is fairly high ($\nu \approx 10^7$ sec^{-1}) and therefore only long waves are reflected from it; for medium and short waves the D-layer is chiefly an absorber. It is present only in daytime.

The E-layer lies at a height of 100–120 km; the height of its base remains very constant through the day and year. The maximum electron concentration inside it is about $1-2 \cdot 10^5$ el/cm^3 in daytime and drops to $5 \cdot 10^3$ el/cm^3 at night; consequently, at night it is usually not detected with the radio-wave frequencies used for ionospheric sounding. The collision frequency in this layer is $\nu \approx 2 \cdot 10^5$ sec^{-1}.

The F-layer has a more complex structure and is fairly diffuse vertically. A distinctive feature is that it splits into two layers (F_1, F_2) in daytime in the summer season. The F_1-layer lies at a height of about 200 km and the F_2-layer at about 250–350 km and over. The maximum

electron concentration in the F-layer is of the order of $2 \cdot 10^6$ el/cm^3, sometimes reaching values of about $5 \cdot 10^6$ el/cm^3 in daytime. The number of collisions is considerably lower than in the E-layer, and the F-layer plays a very important role in the propagation of short radio waves.

Aside from these stable layers, fairly long-lasting (they appear for several hours) ionized formations acting as additional reflecting layers are observed comparatively frequently; they are called sporadic layers. Of these the best studied and most regular is the sporadic E_s-layer, which lies at the same height as the E-layer but has a higher electron concentration (up to about $1 \cdot 10^6$ el/cm^3). When it is present individual clouds of high ionization, as it were, appear in the E-layer. Also observed is the sporadic F_s-layer, encountered in winter at night.

Detailed studies have established that the ionosphere is extremely unstable and nonuniform. The reflecting layers often consists of a series of individual ionized formations in constant motion. These ionospheric clouds, as they are called, form and fade continuously, and are themselves made up of series of even smaller inhomogeneities due to local thickening and thinning of the electron density. The size of these inhomogeneities varies, from a few tens to several hundred meters (frequent size about 200–300 m). By observing the motion of all these inhomogeneities it is possible to draw conclusions concerning wind speeds in the ionosphere; values obtained vary between 20–30 and 400–500 m/sec with a most frequent value of about 90 m/sec (slightly higher in the F-layer and lower in the E-layer).

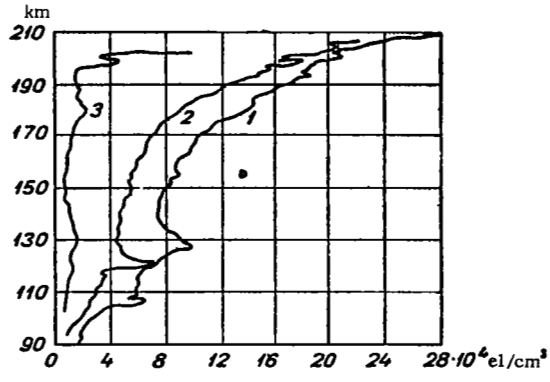


FIGURE 188. Distribution of electron concentration, according to rocket data 1–06:18, 16 May 1957; 2–06:27, 25 August 1957; 3–19:54, 19 September 1957.

The data presented here characterize the average state of the ionosphere encountered in the majority of cases — its "normal" state. However, from time to time this state is disrupted. Particularly severe disruptions are observed during so-called ionospheric storms, which are usually accompanied by magnetic storms and occur with greatest severity and frequency in polar latitudes. During ionospheric storms the state of the ionosphere becomes extremely unstable and the ionosphere itself assumes a pronounced cloudy structure, especially in the F-layer region; in addition, radio communications are severely disrupted, especially in the high

frequencies. The electron concentrations of the E- and F-layers rise sharply during storms, resulting in strong absorption of reflected radio waves in these layers.

Ionospheric storms usually appear unexpectedly and last a fairly long time (sometimes several days). After they are gone the ionosphere slowly and gradually returns to its usual state. The appearance of storms may be attributed to the ionization which occurs when corpuscular fluxes from the sun invade the atmosphere.

In another form of disruption of the ionosphere, a sharp increase in the absorption of all short radio waves occurs unexpectedly over the portion of the globe illuminated by the sun's rays, sometimes leading to complete cessation of radio communication in these wavelengths. Such disruptions are usually observed at the same time as chromospheric flares on the solar disk, during which the sun releases an intensive short-wave radiation which penetrates deeply into the atmosphere.

The close link between the state of the ionosphere and solar activity is also expressed in the fact that when the latter increases the ionization of the entire ionosphere increases, and regular variations with the 11-year cycle can be observed.

The complexity of the structure of the ionosphere and of the phenomena taking place inside it is evident from what we said. Exhaustive interpretation of these questions will require further observations. In the last few years new data have been obtained with the aid of artificial satellites and geophysical rockets. Determinations of the electron concentration, either by the measurement of radio wave dispersion (recording the phase difference between two radio waves with different frequencies) or by the measurement of the rotation of the plane of polarization of radio waves (Faraday effect), have shown that the ionosphere does not possess sharply demarcated layers. Ionization grades from one region of the ionosphere to the other without sharp changes. There are only local maxima, superposed on a general monotonic increase in the electron concentration to the principal maximum in the F_2-layer, as indicated in Figure 188.

The notion of the ionosphere as a system of clear-cut ionospheric layers is now being reconsidered.

Chapter 30

THE ELECTRIC FIELD IN THE ATMOSPHERE

§ 1. Fundamental relations

As a result of the joint action of the charge on the earth's surface and the space charges contained in the atmosphere, an electric field is always present in the atmosphere. This field can have different directions at different points in the atmosphere and can vary over a very broad range. The vertical component of the field strength is almost always considerably larger than its horizontal components and is usually directed downwards towards the earth's surface, i.e., as if the latter were negatively charged; it is customary to describe this field strength as positive.

The electric field of the atmosphere, like any electrostatic field, can be characterized at any point by the potential V. Having determined the value of V at a certain point with the coordinates x, y and z, we have the following expressions for the components of the field strength along the coordinate axes:

$$E_x = -\frac{\partial V}{\partial x}, \quad E_y = -\frac{\partial V}{\partial y}, \quad E_z = -\frac{\partial V}{\partial z}; \tag{1}$$

the total field strength will be

$$E = \sqrt{E_x^2 + E_y^2 + E_z^2}. \tag{2}$$

By drawing continuous surfaces through points with the same value of the potential, we obtain level or equipotential surfaces. According to the universally known properties of the potential, these surfaces can neither intersect nor touch. Taking at any point of the equipotential surface the derivative of the potential along the normal (n) to it, we obtain the total field strength at this point

$$E = -\frac{dV}{dh} = -\operatorname{grad} V. \tag{3}$$

FIGURE 189. Distribution of level surfaces

From this it follows that the field strength will be greater in magnitude at places where the equipotential surfaces converge, and vice versa. Equipotential surfaces roughly follow the contours of the earth's surface, converging over all prominences and diverging over hollows. Their distribution gradually smooths out with height, as shown schematically in Figure 189.

Immediately at the earth's surface, which we are taking to be a conductor,

the field strength is given by

$$E_0 = -\frac{dV}{dh} = 4\pi\sigma, \tag{4}$$

where σ is the density of surface charge at the given point.

In the presence of space charges in the atmosphere, the following relation (Poisson's equation) must also be satisfied at every point of the atmosphere

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -4\pi\rho, \tag{5}$$

where ρ is the density of space charge.

From (3) it is evident that the field strength at any point is numerically equal to the change in potential per unit distance along the normal to the equipotential surface in the direction of decreasing potential (its gradient). In practice the potential is measured in volts and the distance in meters; thus the field strength and the potential gradient (which is equal to it but opposite in sign) are measured in volts per meter (V/m).

The above considerations may be applied to the atmosphere, but only with a certain approximation. Indeed, though the atmosphere is an inhomogeneous medium in the electrical sense, in most cases it is a great deal less inhomogeneous in the horizontal than in the vertical direction. This makes it possible to disregard the horizontal inhomogeneities in the first approximation and treat the field over an extensive flat plain as homogeneous. In describing the field in the atmosphere one can simplify equations (3) and (5) by replacing the normal and the z coordinate by the height h and taking $\frac{\partial^2 V}{\partial x^2} = \frac{\partial^2 V}{\partial y^2} = 0$. Then

$$\frac{d^2 V}{dh^2} = -\frac{dE}{dh} = -4\pi\rho. \tag{6}$$

Observations at the surface give on the average a value of about 130 V/m for the potential gradient. The mean value of the surface charge density of the earth σ can be calculated from equation (4); it proves to be

$$\sigma = -3.45 \cdot 10^{-4}\ \text{CGSE}, \tag{7}$$

which corresponds roughly to $7 \cdot 10^5$ elementary charges per square centimeter. Since the surface of the entire earth is $4\pi a^2$, where a is the earth's radius, the earth's charge must be $Q = 4\pi a^2\sigma \approx -10 \cdot 10^{14}\ \text{CGSE} = -5.7 \cdot 10^5\ \text{C}$.

§ 2. Electric field strength according to observational data

A great many different methods are employed to measure the electric field. The most important are the collector methods and the plate method. In the former the values of the potential are measured at two points situated at different heights with the help of so-called collectors and the potential gradient is determined from the ratio $\frac{V_2 - V_1}{h_2 - h_1}$. In the second method, which makes use of the relation $E = 4\pi\sigma$, σ is measured and the field strength is determined directly from its value. Observations performed frequently in

a field deformed by relief usually [do not] lead to values corresponding to a plain; for this reason observed values are multiplied by the so-called reduction factor. Details on methods of field measurement will be found in the appropriate manuals.

Measurements of the electric field near the earth's surface reveal a high degree of variability depending on meteorological phenomena, especially precipitation, cloud, thunder, etc.; in their presence the potential gradient may fluctuate within several tens of volts per meter and often changes direction (sign). These irregular fluctuations can be regarded as disruptions of a certain "normal" field observed in cloudless weather. By a careful selection of the data of observations carried out on such days, one can identify the characteristics of this normal fine-weather field and establish its regularities.

Observations show that the potential gradient is usually positive over both land and sea, i. e., the atmosphere is positively charged with respect to the earth's surface; further, as indicated earlier, its mean value can be taken to be 130 V/m. Values at individual points may differ markedly from this figure (Table 93).

TABLE 93

Potential gradient at various points

Observation point	$\frac{dV}{dh}$	Observation point	$\frac{dV}{dh}$
Pavlovsk (Leningrad)	171	Yuzhno-Sakhalinsk	147
Potsdam	203	Tashkent	128
Kew (London)	363	Samoa	115
Uppsala (Sweden)	70	Java	86
Spitsbergen	78	Antarctica (Mirny), after IGY data	140
Sverdlovsk	161	Oceans, after IGY data	134
Irkutsk	192		

Remark. Values of the potential gradient vary with time of observation and during the performance of a set of observations.

On the average the potential gradient is slightly greater in middle latitudes, decreasing towards both equator and poles.

As one rises in the atmosphere the potential itself increases but the gradient decreases very rapidly; this is seen from the following (averaged) data:

Height (km)	0	0,5	1,5	3	6	9
$\frac{dV}{dh}$ (V/m)	130	50	30	20	10	5

This falling off of the potential gradient is due to the presence of space charges in the atmosphere. If the vertical distribution of the gradient is known, the distribution of the space charge density ρ can be determined from formula (6). Results obtained in this manner for individual layers are presented in Table 94.

The data given in the last row of Table 94 show that about $7 \cdot 10^5$ elementary charges are contained in a column up to 9 km high; this corresponds approximately to the surface density of charge at the earth's surface. Thus the space charges compensate, as it were, for the earth's charge.

Space charges form in the atmosphere as a result of a wide variety of processes, for example the uneven displacement under the influence of the electric field of ions carrying different signs of charge. Space charges may appear in various processes of electrification, friction, formation of

TABLE 94

Vertical distribution of space charge density

Thickness of layer (km) of 1 cm^2 cross section	Space charge of entire column (elementary charges)	Mean density ρ (elementary charges per cm^3)
0–0.5	$4.45 \cdot 10^5$	8.9
0.5–1.5	$1.11 \cdot 10^5$	1.1
1.5–3.0	$0.56 \cdot 10^5$	0.38
3.0–6	$0.55 \cdot 10^{-5}$	0.18
6–9	$0.28 \cdot 10^{-5}$	0.09
0–9	$6.95 \cdot 10^5$	0.77

water spray, etc., when charged particles of predominantly one sign are released into the atmosphere and subsequently scattered over considerable distances from their place of origin (dust, smoke, fragments of ice crystals, etc.). Measurements carried out by various special methods show that space charges vary greatly with place and time; particularly large values (up to 25–30 CGSE m^3) are reached during thunderstorms and precipitation.

The vertical variation of E, λ and i under fine-weather conditions is illustrated schematically in Figure 190.

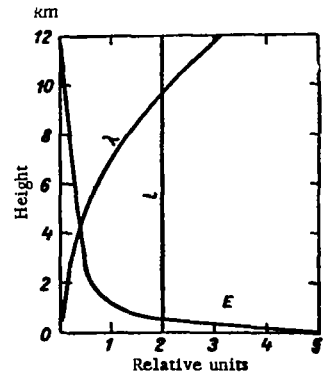


FIGURE 190. Vertical distribution of electric field strength E, conductivity λ and vertical current i (schematic)

The field strength drops off particularly sharply with height near the earth's surface. Up to a certain height (of the order of 50–100 m) which depends on the field strength in the atmosphere, on the degree of ionization and on turbulence, this lower layer receives a supply of ions carrying

charges opposite to the earth's charge, i. e., usually positive ions; owing to the well-known electrode effect the unipolarity of the ions $\frac{n_+}{n_-}$ and the conductivity $\frac{\lambda_+}{\lambda_-}$, as we saw earlier, are greater than unity at the surface.

In reality the vertical variation of E is considerably more complicated. The smooth decrease is disrupted by the influence of space charges (chiefly of clouds), the presence of layers of low conductivity, and so on; as a result there may sometimes even be an increase in the field with height. This question will be considered later.

§ 3. Daily and annual march of the electric field strength

The electric field strength displays regular annual and daily oscillations The annual march of the field in the middle latitudes of the northern hemisphere is simple—maximum in winter (December–February) and minimum in summer (June–August); the amplitude of the annual oscillation decreases away from the middle latitudes toward southern and polar regions. A

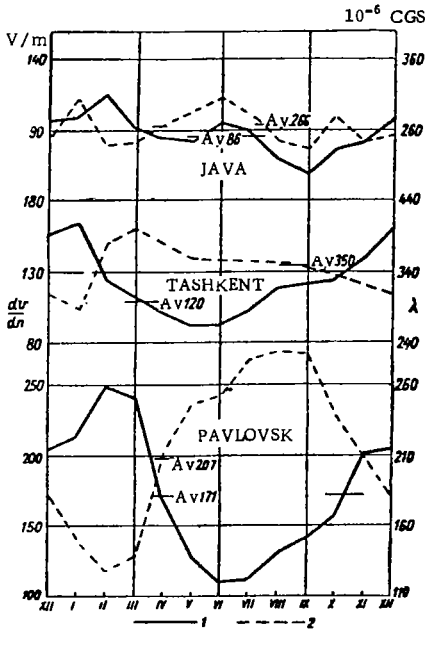


FIGURE 191. Annual march of field strength E (1) and conductivity λ (2)

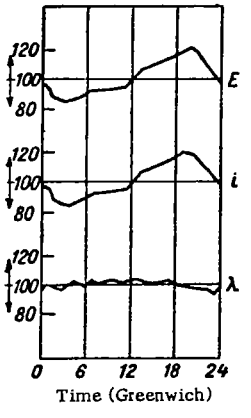


FIGURE 192. Daily march of field strength E, conductivity λ and vertical current i over oceans

typical annual march is that observed, for instance, at Pavlovsk and Tashkent (Figure 191). In tropical regions, e. g., Java, a periodicity with two maxima and two minima is more frequent.

In the southern hemisphere the lowest and highest values occur on the average in the same months as in the northern hemisphere (though with a smaller amplitude); thus the variation of $\frac{dV}{dh}$ over the year proceeds in the same way for the entire earth.

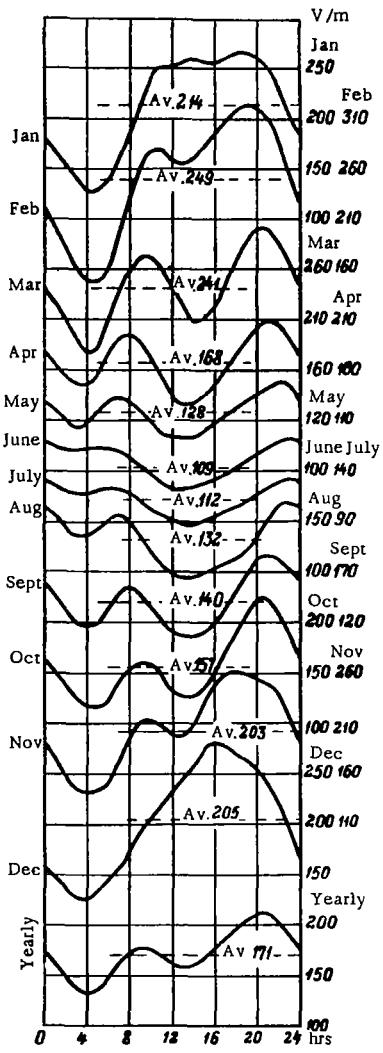


FIGURE 193. Daily march of the field strength at Pavlovsk

The daily march of the field strength is particularly interesting. Disregarding purely local secondary divergences, the following basic features may be mentioned.

1. The daily march is simplest and most regular over oceans, where it remains nearly unchanged through the year, with a maximum at about 18 hrs and minimum at 4 hrs Greenwich time. Extremum values occur simultaneously over all oceans (Figure 192). These changes, which take place at a certain universal time, have been called unitary variations.

2. A daily march of the same form is observed in polar regions and globally at certain heights above the level of the earth's surface; in the winter months it is frequently observed at the surface as well, but only in high latitudes.

3. Overland near the earth's surface the daily periodicity is more complicated. In most places it is in the form of a double wave with two maxima and two minima, the secondary maximum and minimum occurring at local times. The principal minimum is observed in the early morning (3–5 hrs) and the secondary minimum in the afternoon; the principal maximum occurs in the evening (18–22 hrs) and the secondary maximum before noon. The times of occurrence of the daily extrema shift from month to month, and in general the shape of the daily curve varies strongly over the year. In summer the afternoon minimum is particularly deep.

Detailed analysis of the features of the daily march shows that the appearance of secondary maxima and minima is due to the influence of the surface layers and that the basic type of daily march is the simple wave with minimum in the morning and maximum in the evening; the latter can in fact be detected everywhere by appropriate processing of the observations.

Thus the overland daily march of the field strength (in the surface layers) can be pictured as due to the superposition upon the principal simple march, observed in fairly pure form over oceans and due to causes common to the

entire globe, of those fluctuations of the field which stem from variations in conductivity and the influence of space charges and which sould be termed local.

Depending on which of the indicated causes is dominant at a certain instant at a given spot, the character of the daily march also changes.

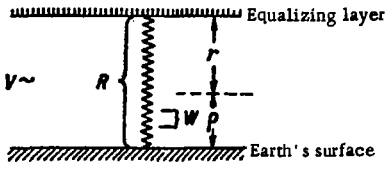


FIGURE 194. Scheme of electric field in atmosphere

For a more detailed consideration of this question, let us turn to the diagram in Figure 194. In this figure the electric field is treated schematically as the field of a plane condenser, one of the plates of which is the earth's surface while the other is the upper atmospheric layers (equalizing layer); the conductivity of these is very high and therefore the field experiences a large "shrinkage" Let the potential difference between the plates be V. We denote the total resistance of the atmospheric column of unit cross section between the plates by

$$R=\int_0^H r\,dh=\int_0^H \frac{1}{\lambda}\,dh,$$

where H is the distance between the plates and $r=\frac{1}{\lambda}$ is the resistivity of the atmosphere at the point under consideration.

A current will then flow between the plates; its density can be written as

$$i=\frac{V}{R}. \tag{8}$$

On the other hand, the conduction current at a certain point in the atmosphere where the field strength is E (in the absence of other sources of current) is given by

$$i_{cond}=E\lambda. \tag{9}$$

Equating the two last expressions, we obtain

$$E\lambda=\frac{V}{R}. \tag{10}$$

Taking the logarithmic derivative with respect to time, we find that

$$\frac{1}{E}\frac{dE}{dt}=\frac{1}{V}\frac{dV}{dt}-\frac{1}{\lambda}\frac{d\lambda}{dt}-\frac{1}{R}\frac{dR}{dt}. \tag{11}$$

The above makes it possible to analyze the causes which lead to variations in the field strength. Indeed, consider the case where two out of the three quantities V, R and λ, namely R and λ, remain constant in time; then, as is easily seen, the variations of the field strength E will be determined by the variations of the potential difference V (a question we will consider below), and since the latter spread very rapidly (within a few

seconds) over the entire globe and take place practically simultaneously for the whole earth, the variations of E which they induce will occur simultaneously everywhere.

This gives rise to the unitary type of field variation mentioned above in which the variations of E take place simultaneously and identically throughout the earth's surface and are proportional to the variations of V. In the general case both R and λ may vary. In the first approximation the resistance of the entire column of air can be considered constant throughout the day in a certain area; in general its influence on the regular diurnal variations of the field will therefore be insignificant and can be disregarded. Changes in its value from place to place can explain the differences in the values of the field strength at different places (E is smaller where R is greater, and vice versa). However, if λ varies over a certain limited area of the atmosphere*, then for constant R and V the field strength will change in the opposite direction. Thus according to this scheme the variations of E will be determined by the variations of V and λ. But the scheme does not allow for the influence on E of the space charges which are always present in the atmosphere, and the amount and distribution of the latter vary in time. In reality the total change in the field strength at any point ΔE is determined by $(\Delta E)_V$, which depends on the variation of V, and also by $(\Delta E)_\lambda$ and $(\Delta E)_\rho$, which are governed by the variation of λ and ρ; thus we can write

$$\Delta E = (\Delta E)_V + (\Delta E)_\lambda + (\Delta E)_\rho. \qquad (12)$$

Obviously, the roles of these factors may be very different in different places. Table 95 gives their values in percents of the total variation of the field strength.

TABLE 95

Values (%) of the individual factors determining the diurnal variations of the field strength

Observation point	ΔE %	$(\Delta E)_V$	$(\Delta E)_\lambda$	$(\Delta E)_\rho$
Pavlovsk (Leningrad)	100	26	41	33
Tashkent	100	17	56	27
Alaska	100	63	29	17

Many special features characteristic of the variations of the field strength and its distribution over the earth's surface can be explained qualitatively by applying similar arguments.

§ 4. Correlations between the electric field strength and other electrical and meteorological elements of the atmosphere

The pattern of variation of the electric field strength described above shows up fairly clearly during observations on "normal" days. It can also

* Which would have no marked influence on R.

be discerned in average results by processing observations over a long period (several years). Usually, however, the electric field experiences irregular and often large fluctuations due to the influence of space charges and to variation of the conductivity. Therefore a very close link exists between the field strength and other electrical and meteorological characteristics of the atmosphere.

This is already shown by the fact that the mean values of $\frac{dV}{dh}$ are markedly different in different air masses. Table 96 gives values of the latter after Koenigsfeld.

TABLE 96

Air	$\frac{dV}{dh}$ (V/m)
Arctic	282
Polar Continental	201
Continental	159
Polar and Maritime	117
Maritime	79

The link between $\frac{dV}{dh}$ and the conductivity λ is particularly close: these quantities vary in opposite senses not only in the daily and annual marches (cf. Figure 191) but also at each individual instant.

Owing to the close link between the field strength and the conductivity, all meteorological conditions affecting the latter also influence the former — and sometimes even more conspicuously so. The dust content has a particularly strong influence; in dust-filled air the conductivity drops and the field strength increases sharply. Naturally, the link between the field strength and the number of uncharged nuclei is also close, the former increasing strongly when the latter increases. The connection with visibility is obvious: when it improves $\frac{dV}{dh}$ decreases strongly.

As in the case of the conducitivity, the correlation with individual meteorological elements is only indirect and depends on the degree to which the given meteorological conditions tend to induce changes in space charges and their distribution in the atmosphere, and also changes in conductivity.

The link between the potential gradient and buoyancy of water vapor is particularly interesting: as the latter increases $\frac{dV}{dh}$ also increases. This relationship is reflected in the similarity between their diurnal periodicities.

In a number of cases a link has been established between the field strength and the wind. At sufficiently high wind speeds charged particles of dust are torn away from the surface and the resulting space charges exert an influence on the field. Sometimes the wind carries space charges laterally from source regions (large cities, industrial zones, etc.). The field strength changes sharply and describes irregular fluctuations in fogs, in the presence of clouds and especially during precipitation and thunderstorms. Since conductivity in fog and cloud is significantly (three times on the average) lower, in the presence of a thin layer of fog or cloud the

field strength inside it will increase accordingly. At the same time, space charges formed in such layers under definite conditions also produce a certain effect. This question will be considered in detail in the next chapter.

It may be assumed that high clouds have almost no appreciable influence on the field strength near the surface. The influence of low clouds can be considerable, especially when large particles are formed inside them and precipitation is released. Since the space charges which appear can be both positive and negative and may be of considerable magnitude; the field can change within a very broad range, sometimes even reversing direction. Obviously, clouds and space charges in the atmosphere distort the vertical distribution of the field described above, and moreover do so extremely irregularly.

As a result of the influence of such purely local conditions, the field strength and its daily march at the earth's surface may frequently be somewhat different even at two neighboring points.

Chapter 31

THUNDERSTORM ELECTRICITY

§ 1. Electric charges of precipitation

It has long been established that all types of particles of precipitation — rain, snow, hail, etc. — are electrically charged. The charge carried by particles of precipitation is measured by collecting them in a suitably placed insulated container connected to an electrometer. This makes it possible to determine the density of the current induced by the transfer of charged precipitation particles (convective current of precipitation) and the charge per unit volume (mass) of fallen water or per droplet (by capturing the drops individually).

Such observations have led to a number of definite conclusions, the most significant of which are the following.

1. During any one rainfall (snowfall) a mixture of positively-charged, negatively-charged and also uncharged droplets (snowflakes) is released. The number of positively-charged droplets usually exceeds the number of negatively-charged ones, on the average by a factor of roughly 1.5; the ratio varies from 1:1 to 30:1 in individual cases. At the same time the average negative charge carried by an individual drop is greater than the corresponding positive one.

2. The charges on individual droplets vary a great deal, averaging about $1.1 \cdot 10^{-12}$ C for positively-charged droplets and $1.3 \cdot 10^{-12}$ C for negatively-charged ones. Charges on hailstones are particularly large. While droplets of steady rain are charged on the average to a potential of 0.5 to 10 V, the potential on thundershower drops reaches up to 300 V with a mean value of about 40 V.

3. Different types of precipitation differ in electrical characteristics; in general, larger drops also carry large charges; particularly large charges are carried by showers (in thunderstorms) and hail. The smallest charges are carried by steady rains.

In rain (hail) showers the current density may reach 10^{-11} A/cm^2, while in steady, light rains it amounts to roughly 10^{-15} A/cm^2 and less.

Table 97 gives the mean charges of droplets and the current densities for different kinds of precipitation.

4. The sum total of the positive charges carried by precipitation, and also the sum of the time intervals over which positively-charged precipitation is released, is greater than for negative charges.

5. At different heights in the same rainfall zone charges on droplets may differ in both magnitude and sign. Together with other indirect data, this leads to the conclusion that: a) drops captured at the surface fall from different cloud parts in which different processes lead to the formation of their initial charge;

b) the charges of drops can change during the long path from cloud to surface.

The question of the causes which produce charges on precipitation is extremely complex. On the basis of general physical considerations and numerous laboratory studies, one could point to a whole series of processes, all unquestionably capable of leading to the electrification of droplets and snowflakes followed by the transfer of their charge. However, as conditions under which electrification takes place in clouds are not always known, it is difficult to say what processes are dominant in each individual case and whether they are sufficiently intensive from the quantitative standpoint. The number of existing measurements inside clouds and especially thunderclouds — where processes achieve maximum intensity — is still small, and one is frequently obliged to base an opinion on indirect considerations.

TABLE 97

Particle charges and currents in various forms of precipitation

Precipitation	Q_{av}(C)	Q_{max}(C)	i_{av}(A)	i_{max}(A)
Steady rain	10^{-15}–10^{-14}	$5 \cdot 10^{-13}$	$5 \cdot 10^{-15}$	$5 \cdot 10^{-14}$
Shower	10^{-13}–10^{-12}	$5 \cdot 10^{-12}$	$5 \cdot 10^{-14}$	$5 \cdot 10^{-12}$
Hail	10^{-11}	10^{-10}	1·	$1 \cdot 10^{-11}$
Snow	10^{-12}–10^{-11}	$5 \cdot 10^{-11}$	$1 \cdot 10^{-14}$	$5 \cdot 10^{-13}$

The following processes of cloud electrification are now recognized: 1) the collision of cloud particles polarized in an electric field and the exchange of charge between particles of different sizes; 2) the capture of ions by particles of precipitation; 3) the breaking of water drops and disintegration of ice crystals falling in air; 4) various processes of electrification and separation of charge taking place during transformations of the aggregate state of water clouds (freezing, thawing, etc.) and many other minor processes.

As early as the beginning of this century Elster and Geitel pointed out that drops falling in an electric field were polarized. As a result charges of one sign (negative for a normal electric field in the atmosphere) are concentrated in the upper half of the drop and charges of the opposite sign and equal magnitude in the lower half. Larger drops collide with minute droplets when falling but, as experiments have shown, not every collision leads to fusion. However, electrical contact with the exchange of charge does occur in such collisions, as a result of which the droplets may acquire proper charges of different signs depending on the place where they break off one from the other. This induction theory, as it is called, though not sufficiently general, can account for the electrification of droplets of precipitation in a number of particular cases (gentle, steady rainfall).

Wilson's theory, which considers the capture of ions from the air by particles of precipitation polarized in an electric field, merits closer examination. Depending on the ratio between the rate of fall of the drop and ion velocity in the electric field, the drop will capture predominantly negative or positive ions; in most cases the ions captured are negative.

This theory, subsequently developed by a number of researchers and verified by numerous laboratory experiments, leads to quantitative agreement with the charges observed in reality. There is no question but that in many cases the process indicated by Wilson is fundamental to the electrification of particles of precipitation.

Furthermore, a particle which encounters positive and negative ions during its fall may acquire charge owing to simply to the capture of ions on their surface (adsorption theory). Thus Ya.I. Frenkel' pointed out that *minute cloud droplets as liquid dipoles adsorb negative ions more easily* and acquire a certain negative equilibrium charge, given by $r\zeta$, where r is the droplet radius and ζ the so-called electrokinetic potential (for water $\zeta = -0.25$ V). Subsequently the sign and magnitude of the droplet charge are controlled by the processes of diffusion of ions to the droplet surface and therefore by the concentration of ions, positive and negative, along the fall path of the droplet, as well as by charge transfer in droplet coalescence.

The balloelectric phenomena, a term applied, in particular, to the phenomenon of electrification of droplets on breakup (Lenard effect) have *considerable importance for charge formation on precipitation. In the* Lenard effect, when a sufficiently large drop breaks up into fragments while falling in the atmosphere, the large fragments acquire a certain charge (usually positive) and the small fragments an opposite charge. An analogous phenomenon takes place in the fragmentation of solid particles, and also in the disruption of waves in sea surf, near waterfalls and mountain torrents, etc. It is caused by the disruption of the electrical double layer at the surface of the drops. However, this effect can appear only in large drops ($r \sim 2-3$ mm) and for large updraughts.

Processes of electrification in clouds containing supercooled droplets and ice crystals are particularly important for charge formation on particles of precipitation. They all involve phase transitions of water and display great diversity. Experiments by Workman and Reynolds have shown that when dilute solutions of various salts freeze large potential differences, reaching several tens of volts (i. e., much greater than the ordinary contact potential) appear at the ice-water boundary. A possible explanation for this is that ions of the solution, positive or negative depending on the conditions of the experiment, freeze into the ice. Charge formation is also observed in the thawing of ice, sublimation, freezing of supercooled drops and many other phenomena. This has been confirmed by numerous experiments, which have shown that the charges formed may be very considerable; in thunderclouds they may lead within a few minutes to the formation of charges of the order of coulombs, sufficient for the development of lightning.

Certain of the theories of charge formation and separation considered above do not always agree with known facts concerning the meteorological and electrical structure of thunder clouds, while certain others are open to criticism on quantitative grounds. Mechanisms involving the creation of a solid phase inside thunder clouds and in particular particles of graupel, are clearly to be preferred.

§ 2. Distribution of charges in thunder clouds

The processes of electrification considered above lead to the formation of equal amounts of positive and negative charges, and it is only upon their spatial separation that substantial concentrations of space charge arise in different parts of the cloud. This separation of charges takes place chiefly under the influence of gravity, as well as that of electric fields and various other factors. However, electric fields appear both inside and outside the cloud when separated space charges are formed and can strongly distort the normal field of the atmosphere. Changes in the magnitude or distribution of these space charges in the cloud are accompanied by the sharp changes in the field which can usually be observed during thunderstorms and showers at the earth's surface.

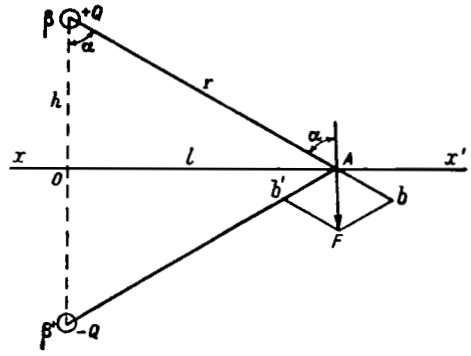


FIGURE 195

Let us consider this important question in greater detail. Let us assume that the charge Q occurs in a small spherical volume at the point B, situated at the height h above the conducting plane horizontal surface of the earth xx' (Figure 195). In calculating the field due to this charge at the point A we will use the method of images and assume that the field at A is the same as it would be in the absence of the conducting surface xx' if a charge $-Q$ were acting at the point B', which is the mirror image of B with respect to the plane xx'. But the field due to the charge $+Q$ is numerically given by $\frac{Q}{r^2}$ and is represented by the vector $\vec{Ab}$. The field due to the charge $-Q$ lying at the point B' is given by $-\frac{Q}{r^2}$ and is represented by the vector $\vec{Ab'}$. Summation of these two fields, the horizontal components of which cancel each other, gives the following for the field strength at the point A:

$$E=\frac{2Q}{r^2}\cos\alpha \qquad (1)$$

or, since $\alpha=\frac{h}{r}$ and $r^2=l^2+h^2$,

$$E=\frac{2Qh}{r^3}=\frac{2Qh}{(l^2+h^2)^{3/2}}. \qquad (2)$$

Obviously, the influence of the space charge will be strongest at the point Q directly beneath the cloud ($l=0$) and will decrease rapidly as one moves away from this point; thus, for instance, for $Q = 10$ C at the height $h = 2$ km we find:

l (km)	0	2	5	8	15
$\frac{dV}{dh}$ (V/m)	45 000	15 900	2300	620	95

When the charge changes instantaneously by the amount ΔQ, as for instance, in lightning, the field strength also changes, by an amount which is obviously given by

$$\Delta E = \frac{2h\Delta Q}{r^3} = \frac{2h\Delta Q}{(l^2+h^2)^{3/2}}. \qquad (3)$$

The field will also change if there is a displacement of charge in the atmosphere (if r changes).

Usually clouds contain space charges which are opposite in sign and possibly different numerically in different parts of the cloud. To calculate the field strength set up in this way one can make use of the method of super-superposition for the fields set up by each charge individually.

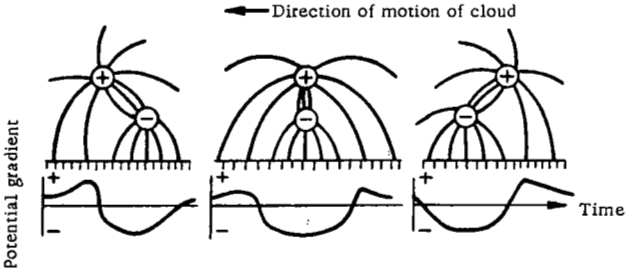


FIGURE 196. Change in electric field upon movement of cloud (dipole)

Figure 196 illustrates the distribution of the potential gradient at the earth's surface upon displacement of a 'dipole' cloud (schematic).

By observing the field distribution over the earth's surface at various distances from the cloud or following alterations of the field at a given point upon passage of the cloud, one can draw certain conclusions concerning the charge distribution inside the cloud. However, these conclusions will not be entirely reliable, as they will require a number of assumptions regarding the structure of the cloud; direct measurement remains the only reliable method for studying the charge distribution in a cloud.

As long as the formation and release of precipitation do not take place, space charges in the cloud remain small, the cloud as a whole is essentially neutral and the electric field is small both inside and outside. When the larger particles of precipitation are formed there is a separation of variously charged particles (large and small) and considerable space charges appear in the cloud. In view of the wide variety of processes responsible for the formation of charges on particles and their subsequent separation, the distribution of space charge in clouds is naturally highly variable. Especially large charges are formed in shower and thunder clouds.

Observations show that in these clouds negative charges are concentrated predominantly in the lower part of the cloud up to the level of the 0° isotherm and positive charges above the −12° isotherm; a mixture of positive and negative charges forms in the intermediate layer (between the 0 and −12° isotherms).

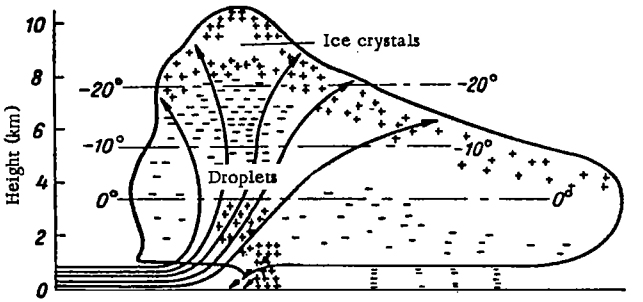


FIGURE 197. Charge distribution in thunder cloud (after Simpson and Scrase)

In well-developed thunder clouds limited regions with positive charge are occasionally encountered in the lower, negatively-charged part. This distribution (shown schematically in Figure 197) is the result of two distinct physical processes, one acting in the upper, glaciated part of the cloud (above the −12° isotherm) and the other in the lower part (below the 0° isotherm). In the upper regions owing to processes of electrification of the ice formations the small ice splinters receive a positive charge and concentrate at the top while the larger, negatively-charged crystals drop into the lower regions. Here temperatures are above zero and the crystals melt, forming droplets of water; in the presence of strong turbulent motions in a limited region of the cloud, these droplets may be broken up mechanically. As a result (balloelectric effect) the large drops assume positive charges. They form regions of positive charge in the lower part of the cloud. Such regions, however, can appear only in fully developed thunder clouds and in the presence inside them of particularly strong updraughts.

It should be emphasized that the above scheme represents only the most general aspects of the charge distribution in thunder clouds; the actual pattern may often depart significantly from this scheme.

Measurements of the electric field inside thunder clouds were first carried out by Simpson and his colleagues. They used sounding balloons equipped with instruments based on the measurement of the corona discharge current from a point (altielectrograph).

Simpson and Robinson constructed a theoretical model of a thunder cloud which agrees with the observed heights of charge centers and the observed surface values of the field. The agreement between the theoretical and experimental distributions of the field was good. However, only qualitative agreement was obtained. Whereas from the theoretical model it follows that electric fields in thunderclouds can exceed 1000 V/cm, the altielectrograph records indicated that they did not exceed 100 V/cm. Subsequently an error was discovered in the graduation of the altielectrograph. The influence of wind on the intensity of corona discharge from a point had not been taken

into account during graduation. Re-interpretation of earlier results obtained with altielectrographs by Simpson, Chapman and others was attempted with corrections for wind.

The corrected results proved to be one order higher and agree basically with later results obtained with the aid of aircraft investigations of the electric fields in thunder clouds.

§ 3. Thunderstorm discharges. Lightning and the lightning mechanism

The formation of large space charges in a thunder cloud gives rise to spark discharges between individual cloud parts or between the cloud and the ground; these are observed as lightning.

Lightning varies a great deal in appearance. Its principal forms are well known.

The form most often observed is the so-called streak lightning, which has many subtypes; among these zigzag and forked lightning are especially common, while ribbon and rocket-shaped types are less frequent. The mean length of the visible portion of the streak amounts in most cases to about 2–3 km. Estimating the length of lightning streaks inside clouds is very difficult but there are reports of cases where the length of the streak between clouds reached 15–20 km or more (up to 40 km). Lightning moves along a channel with a diameter which reaches up to 40 cm in certain cases but may be estimated as 16 cm on the average.

Another form is sheet lightning, a discharge covering a considerable section of the cloud; in this form a considerable thickness of the clouds virtually flares up for an instant. This form of lightning is distinct from the sky illumination due to a distant, not directly visible streak lightning, which is known as summer lightning.

Beaded and ball lightning are far less frequent. They usually occur after intensive spark discharges.

Beaded lightning consists of several (20–30) small spheres (diameter of about 10 cm) distributed along the discharge channel, which gives it the appearance of beads; the interval between spheres is usually a few centimeters. This phenomenon is very rare and short-lived (about 0.5 sec) and has therefore scarcely been studied at all.

Ball lightning is more frequent than beaded lightning but all observations concerning it tend to be random and basically qualitative. It consists of a luminous ball, sometimes elongated to a pear shape, usually moving slowly (as if floating) in the atmosphere or (sometimes) over the surface of the objects on which it settles.

The diameter of ball lightning at the surface is usually small (about 10–20 cm); however, balls up to several meters (roughly 35 m) in diameter have been reported at a certain height in the atmosphere. In certain cases lightning balls fall out of the thunder cloud, as it were, often exploding with a deafening noise on touching the ground and causing considerable damage.

Owing to the lack of exact instrumental studies and the impossibility of reproducing the phenomenon under laboratory conditions, no reliable explanation of ball lightning has yet been put forward. Among the many extant

hypotheses, one view which is supported by certain laboratory experiments holds that ball lightning represents the site of the combustion of some inflammable gas (hydrogen, propane or other) ignited by spark discharges in the air and therefore is not in the nature of an electrical phenomenon.

Streak lightning, especially between cloud and ground, is the most thoroughly studied of all the forms.

By photographing lightning on a rapidly moving film one can trace its development in time; the current strength in lightning can be determined by collecting lightning by means of special instruments (klydonograph, etc.). Particularly detailed data on the electrical characteristics of lightning are obtained with the aid of specially designed electron oscillographs and also by studying the changes in the cloud's electric field during the development of lightning.

Laboratory research into the mechanism of development of long flashes has been very important for the study of lightning. It should be stressed, however, that while lightning and artificial flashes have much in common, they are not quite identical. This is because lightning is a special kind of discharge (electrodeless discharge) which develops in the absence of conducting (usually metallic) electrodes. In this connection our information concerning the development of lightning between clouds is more limited than that concerning lightning between cloud and ground.

Without delving more deeply into this special question, the complete theory of which has not yet been given, we present the principal results of investigations. Firstly, we note that lightning, although perceived by the eye as a single flash, actually is an interrupted discharge consisting of a series of individual discharges or pulses. The number of pulses is most often 2–3 but in exceptional cases may reach up to 50. The duration of each pulse is about 50–100 μsec and they succeed each other at varying intervals with a most probable value of about 0.03 sec. As a result the total duration of lightning is highly variable depending on the number of individual pulses and the time intervals between successive pulses, and may reach up to 1.5 sec; the most frequent duration is about 0.2 sec. The discontinuous nature of the discharge accounts for the flickering observed in many cases.

Each pulse is itself fairly complex, and consists of two stages:

1) a preliminary and fairly weak discharge, the leader, which usually develops from the cloud to the ground;

2) the main or return stroke, or so-called main channel, which propagates in the reverse direction at a considerably greater speed than the leader stroke and is much stronger.

The development of lightning between a cloud carrying a negative space charge in its lower part and the ground, where positive charges are centered by induction, may be pictured as follows. When the electric field reaches a certain critical value in some part of the cloud, ionization is initiated owing to the fact that the free electrons always present in air acquire high velocities under the influence of the field and ionize the air atoms with which they collide. The electrons formed in this process tend in turn towards the earth and also ionize the air. An avalanche of electrons is formed. As the electrons move they leave behind them positive ions and a large number of excited atoms and molecules. When the latter return to the normal state the emission of photons take place. These photons, catching up with the electron avalanche, in turn produce photoionization and may give rise to the development of new avalanches branching off from the initial

one; most of these develop predominantly in the direction of the ground. Ionization takes place in a narrow channel in which the conductivity is very high (the concentration is up to 10^{13} ions/cm^3). This channel and its branches are gradually filled up with negative charges flowing from the cloud. As the tip of the leader approaches the earth the field strength near it increases; the formation of a positive streamer extending from the earth may sometimes take place.

When the tip of the leader reaches the earth the positive charges accumulated on the ground (ions) and the negative charges in the part of the cloud adjoining the channel and filling the channel become neutralized along the existing conducting path. In this process the channel becomes strongly heated and shines brightly, while the current inside it attains high strength and great instantaneous power. The current persists for about 100μ sec and when the charged portion of the cloud immediately adjoining the lightning channel discharges through the channel the current ceases. But simultaneously with the development of the leader towards the earth, the discharge develops also in the upper layer of the cloud column, penetrating into the higher cloud layers. Therefore after the first pulse has released part of the cloud charge into the ground, another pulse appears; this also begins with the development of a leader, which is again followed by the main stroke. The number of such individual pulses, each of which penetrates more deeply into the cloud, may be as high as 50, as we saw above; all the pulses take the same path already prepared by the first pulse but are less branched than the first.

The development of the leader of the first pulse has one characteristic feature. In addition to being more ramified and considerably stronger than the leader of subsequent pulses, it develops, as it were, by jumps. It traverses a distance of several tens of meters (about 50–60 m), then pauses and after a 30–50μ sec interval again moves a certain distance forward. The impression one receives is that the leader is descending from cloud to ground by steps, and the leader is therefore called a stepped leader. In subsequent pulses the process develops along the path previously prepared by the earlier pulse, and the leader experiences no hindrances; the development times of such leaders are considerably smaller and they are known as dart leaders.

It was found that the effective propagation velocity of stepped leaders, by which is understood the velocity obtained when allowance is made for the time intervals between individual steps, was between $1 \cdot 10^7$ and $2 \cdot 10^8$ cm/sec, with a most probable value of $1.5 \cdot 10^7$ cm/sec. For dart leaders the velocity varies between $1.0 \cdot 10^8$ and $2.3 \cdot 10^9$ cm/sec with a most probable value of about $2 \cdot 10^8$ cm/sec. The propagation velocity of the main stroke is $2.0 \cdot 10^9$–$1.4 \cdot 10^{10}$ cm/sec; the most frequent value is $3.5 \cdot 10^9$ cm/sec.

The development of a lightning flash consisting of three pulses is represented schematically in Figure 198.

In certain cases and especially when tall pointed objects on the ends of which the field strength may be very high are present at the surface, positive stepped leaders may develop from these objects simultaneously with the development of the leader descending from cloud to ground. In such cases the main stroke begins at the instant at which these two opposite leaders come into contact; usually a more prolonged stroke is formed and there are no repeated pulses.

Measurements of lightning currents show that the quantity of charge streaming out during a single lightning may be taken to be 10–50 C with a most typical value of about 20 C; much higher values reaching 150–160 C are observed occasionally. As this amount of charge passes through in a very short time interval, the current strength in lightning is necessarily very high; in certain cases it reaches $(1-1.5)\cdot 10^5$ A and even more. Statistical estimates show, however, that in roughly 80% of the cases the current does not exceed $2\cdot 10^4$ A; on USSR territory, for instance, it amounts to about 10^4 A in 55% of the cases. The direction of lightning currents is such that in discharges into the earth they carry predominantly negative charges to the earth's surface; on the average for the entire earth one can assume that no less than three-fourths of all lightnings carry negative charges.

The path along which lightning develops in the atmosphere at considerable distances from the ground is determined by conditions ensuring easiest development of the leader tip. As the leader descends the earth's surface exerts an increasingly large influence on the lightning path through its surface relief (lightning seeks the highest points where the field strength is greatest) and the electrical properties of the soil (lightning seeks to strike places where the electrical conductivity of the soil is greatest). Depending on which of these two influences is stronger, lightning will strike either at visible high points of the relief or (occasionally) in hollows if the soil conductivity inside them is higher than at adjacent localities.

In certain regions one can single out limited areas of the ground where lightning strikes especially frequently; these are the so-called lightning-prone places.

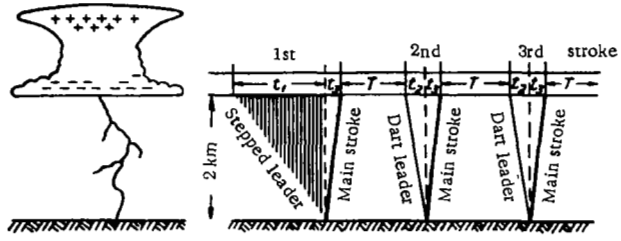


FIGURE 198. Development of lightning between cloud and ground

Development times $t_1 \sim 0.05$ sec; $t_2 \sim 0.001$ sec; $t_3 \sim 0.00005$ sec; $T \sim 0.03$ sec; speed of stepped leader $\sim 1.5\cdot 10^7$ cm/sec; of dart leader $\sim 2\cdot 10^8$ cm/sec; of the main stroke $\sim 3.5\cdot 10^9$ cm/sec.

Special methods are used to protect structures from damage by lightning, including the installation of so-called lightning conductors; the function of the latter is to intercept the lightning current and conduct it into the ground along a pre-arranged path which is harmless to the structure. Pointed rods of sufficient thickness are placed over the area to be protected and carefully grounded. Protection of power transmission lines is sometimes effected by means of grounded metal wires above the current-bearing cables, while particularly important structures are shielded by more complicated devices.

Streak lightning is usually accompanied by a strong booming sound—thunder. Its appearance can be explained as follows: on rapid growth of the lightning current the air temperature inside the channel rises almost instantaneously to a value of the order of $(1.5-2)\cdot 10^4$ degrees. Therefore the pressure in the channel also increases in comparison with the pressure outside it. When the current ceases the temperature in the channel drops equally rapidly and there is a severe compression of the air. These expansions and compressions are of the nature of detonation waves and give rise to sound waves which are perceived as thunder.

The distances over which thunder can be heard usually amount to 20–25 km but are often smaller (average about 15 km). Only in rare, particularly favorable cases can thunder be heard at greater distances. When the thunderstorm discharge occurs in the immediate vicinity of the observer and especially in cases where lightning strikes the ground, thunder takes the form of a strong, sudden shock. As one moves away from the discharge the duration of the thunder increases and characteristic peals are heard. The explanation for these peals is that the sound generated along the long sinuous lightning path though produced simultaneously reach the observer successively; also, the sound is reflected both by objects on the ground and by clouds and surfaces of discontinuity between air masses.

§ 4. Global distribution of thunderstorm activity

The thunderstorm observations carried out at meteorological stations are still confined for the most part to recording the times of occurrence of thunderstorms, occasionally with a subjective point estimate of their intensity. The results of these observations provide a certain general insight into the diurnal and annual periodicity of thunderstorms and their distribution over the earth's surface, the quantitative characteristic of the latter being the number of days with thunderstorms per month, season or year. As these results show, the number of days with thunderstorm is maximum in tropical and equatorial regions and decreases gradually toward the higher latitudes. However, lines of equal thunderstorm incidence are fairly complicated. All regions of maximum thunderstorm incidence usually occur over continents: these are the so-called world thunderstorm foci, at which the number of days with thunderstorm per annum reaches 150–200.

The most intensive world thunderstorm foci include Central Africa, the central part of South America, the south-eastern regions of the USA (Florida peninsula) and South-East Asia.

Over the oceans the Atlantic off the coast of South America and a region in the Pacific Ocean extending from latitude 0° to 25° S. and from longitude 150° E. to 135° W. have frequent thunderstorms.

More detailed maps compiled for individual areas clearly show a dependence of the thunderstorm incidence on local conditions, and reveal so-called local foci of thunderstorm activity. In particular, thunderstorms in mountain regions are more numerous by far than those in low-lying localities, though the latter are more vigorous.

Within the USSR one can point out a number of local centers of thunderstorm activity. For example, thunderstorms are particularly frequent in the Caucasus, in parts of which the number of days with thunderstorm per

year is as much as 60; similar areas can be singled out in the south-eastern part of the European USSR and also in the Ukraine (Kiev and Kharkov areas) as well as in other places.

Statistical estimates show that at every instant thunderstorms cover about 0.36 % of the whole surface of the earth, while the number of thunderstorms per day over the whole earth can be estimated as 44,000; about 100 lightning discharges are formed per second.

At most middle-latitude points of the northern hemisphere the largest number of thunderstorms occurs in summer (June–July) and the smallest number in winter; at that time the rare thunderstorms are feeble and of short duration. At many places, however, thunderstorms are numerous not only in summer (July) but even in winter (January–December); thus a fairly large number of winter thunderstorms is observed in Great Britain, while near Iceland over the ocean the maximum number of thunderstorms always occurs in winter.

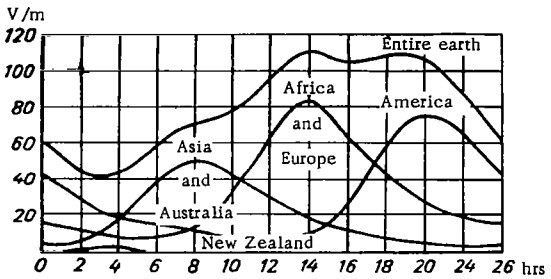


FIGURE 199. Daily march of thunderstorm activity. New Zealand curve covers segment between 120°W. and 150°E.; Asia-Australia curve covers segment between 150° and 60°E.; Africa-Europe curve covers segment between 60°E. and 30°W.; America curve covers segment between 30° and 120°W.

Looking at the distribution over the day one finds that the second half, roughly from 13 to 24 hrs, has more thunderstorms; thunderstorms are most frequent between 15 and 18 hrs, least frequent in the morning (5–7 hrs). Highly significant data on the daily march of thunderstorm activity are given in Figure 199 for extensive sectors of the earth and for the earth as a whole; Greenwich mean times are marked along the abscissa. The curves show that for the entire earth thunderstorm activity displays minimum development at 3–4 hrs and maximum development in the evening (18 – 19 hrs).

§ 5. Silent discharges

During thunderstorms and in cases when the electric field in the atmosphere becomes especially large (snow and dust storms, squalls, hailstorms, etc.) luminous discharges of a special sort are frequently observed at points and sharp corners of objects raised above the earth's surface. These discharges are most widely known as St. Elmo's fire. They occur most

frequently in mountains on sharp projections of rock, buildings, treetops, etc. They are usually accompanied by a characteristic crackling or hissing sound. Their duration is sometimes considerable and frequently reaches several hours. They are observed everywhere and in all seasons of the year; in mountains their frequency is greatest in simmer and in lowlands during snowstorms.

Silent discharges are simply a brush form of corona discharge. This kind of discharge appears when the field strength reaches high values near the electrode (point) owing to the latter's very small radius of curvature but is still not large enough for the formation of a spark discharge. At the initial stage of a corona discharge, when the field strength is still too small for its development, a semi-selfsustaining discharge due to the motion of ions formed under the influence of ordinary atmospheric ionizers takes place near the electrode. This current is extremely small and is completely dependent on the intensity of ion formation; it is not manifested externally in any way. As the field strength rises to a certain critical value which is determined by the form of the electrode and the density of the air the gas suddenly begins to shine with a bluish light nest to the electrode, a characteristic crackling and hissing sound appears and the current at the point increases to values of the order of a microampere or more. The current buildup and glow are a consequence of the ionization and excitation of gas molecules and atoms under the influence of the electrons accelerated by the strong field.

The region of gas glow and impact ionization forms a corona. Its size and the intensity of its glow increase on further growth of the field, and a brush-shaped cone consisting of a series of rapidly moving fine luminous streamers appears about the point against the background of the general feeble glow around the point (brush discharge).

For a positive point charge (positive corona) the glow is distributed almost regularly inside a cone of about 90° at the vertex (of the point end). The electron avalanches develop from the outer edge of the corona towards the point; the fairly immobile positive ions remaining behind the moving electrons produce positive space charges which weaken the field in the immediate vicinity of the point. Due to this the corona appears to be located not directly at the point but on a short luminous column, the corona stem. The length of the streamers in the brush sometimes reaches 15 cm. In negative coronas the development of the streamers begins at the point itself and the electrons travel into the space surrounding the point where they combine with gas molecules, leading to the appearance there of a negative space charge. The corona is considerably smaller than in the previous case and is narrower; the length of the streamers is 2–3 cm. Space charges around the point having the same sign as the point weaken the field and may, if they are not diffused into the surrounding space, lead to the extinction of the corona; therefore the corona current does not remain constant but rather fluctuates continuously and is intermittent in character.

Currents from points are usually observed with the aid of a galvanometer through which a platinum-tipped insulated rod, projecting to a certain height, is earthed.

Under natural conditions the current from the point, the electric field strength E and the wind speed v are seen to be related as follows:

$$i = k(E - C)v,$$

where k and C are constants which depend on the form and dimensions of the point, the sign of the field and certain other parameters.

Currents from a point usually amount to several microamperes. Already the earliest observations revealed that over a considerable time interval the charge conducted to the earth's surface by these currents was predominantly negative (roughly twice as much). Owing to the importance of this question attempts were made to measure the currents flowing, not from artificial points, but rather from the points actually present under real conditions at the earth's surface and in particular trees, vegetation, etc. These complex and naturally highly approximate measurements demonstrated conclusively that the negative charge conducted to the earth's surface by currents was considerably (roughly tens of times) greater than the charge conducted by lightning into the earth.

§ 6. Atmospherics

In § 2 we considered the electrostatic effect of the cloud charge Q on the electric field strengths at the surface. In more general form the relationship between the vertical component of the electric field $E_\perp$ at the surface and the cloud charge Q is expressed by

$$E_\perp = \frac{2hQ}{r^3} + \frac{2}{cr^2}\frac{d}{dt}(hQ) + \frac{1}{c^2 r}\frac{d^2}{dt^2}(hQ), \qquad (4)$$

where h is the height at which the charge Q occurs, r is its distance from the observation point and c is the velocity of light.

The above expression shows that the field $E_\perp$ determined by the influence of the charge Q is made up of three components: 1) the electrostatic field due to the cloud charge (first term); 2) the induction component due to the variation in time of the cloud's electric moment $\frac{d}{dt}(2hQ)$; and 3) the field of electromagnetic radiation, expressed by the third term.

The electrostatic component decreases in inverse proportion to the third power of the distance and the induction component in inverse proportion to its second power; at great distances the dominant factor is the third component, that is the electromagnetic radiation of lightning. Calculations and observations show that the first term remains dominant to distances of the order of $r < 100$ km and that for $r > 500$ km electromagnetic oscillations in the form of short-lived pulses called atmospherics become paramount. Atmospherics are manifested in the radio noises always present in radio receivers. Sometimes these disturbances are so considerable as to prevent reception even of powerful signals.

The study of the electric field changes associated with thunderstorm discharges and especially atmospherics can shed light on the atmospheric processes responsible for their appearance.

Data on the number of incoming atmospherics, their intensity, the direction of arrival, the form of the electric oscillations and so on can be obtained with the aid of special radio-physical methods; by observing these continuously at several reception points and plotting their directions, one can determine the position of the discharge source and track the movements of the responsible storm centers. Such observations are obviously

of considerable practical importance, especially for aviation and radio communications. All sources of atmospherics can be classed into two groups. The first includes the world thunderstorm foci indicated in the previous section. The second group comprises irregular sources appearing only in definite seasons, at locations which are variable and sometimes change very rapidly. Comparison with processes taking place in the atmosphere shows that these atmospherics are for the most part closely related to thunderstorm activity at fronts. This link with fronts and frontal thunderstorms enables an observer to follow the formation and migration of such regions from afar.

Prolonged systematic observations have made it possible to establish the dependence of the number of atmospherics and their intensity on time of day and season. The main features of the daily periodicity are a minimum in the morning, clear-cut at all times and in all places, and a general buildup towards the night hours; the maximum occurs in the middle of the night in the cold half of the year and in the afternoon (15–18 hrs) in summer.

§ 7. Electric current balance in the atmosphere

The electric currents present in the atmosphere vary widely in magnitude, direction and territory covered. They include: 1) the vertical conduction current; 2) the convection current of charge transport by precipitation and aerosols; 3) currents from points at high field strengths; 4) lightning currents to the earth's surface. In addition to these currents, which determine the charge exchange between the earth's surface and atmosphere, other currents are associated with the horizontal transport of space charges by air streams; their density is given by $i_r = v\rho$, where v is the horizontal velocity of the air streams and ρ is the space charge density.

If we consider only the vertical conduction current, which under fine-weather conditions is directed towards the earth, it is easy to show that its action would lead very rapidly to the liquidation of the electric field of the atmosphere. Indeed, since $\frac{dV}{dh} = -4\pi\sigma$, $\frac{d}{dt}\left(\frac{dV}{dh}\right) = -4\pi \frac{d\sigma}{dt}$; but $\frac{d\sigma}{dt} = i_{\text{cond}} = \lambda \frac{dV}{dh}$, and therefore

$$\frac{d}{dt}\left(\frac{dV}{dh}\right) = -4\pi\lambda \frac{dV}{dh}, \tag{5}$$

whence, integrating, we obtain

$$\left(\frac{dV}{dh}\right)_t = \left(\frac{dV}{dh}\right)_0 e^{-4\pi\lambda t}. \tag{6}$$

Numerically for $\lambda = 2 \cdot 10^{-4}$ CGSE we find that after only 30 min $\left(\frac{dV}{dh}\right)_t$ would be reduced to 0.01 $\left(\frac{dV}{dh}\right)_0$.

But the electric field is constantly present and thus we are faced with the fundamental problem of the theory of atmospheric electricity, that of discovering the causes which maintain an electric field in the atmosphere. Since the vertical current (density $i = 2.9 \cdot 10^{-16}$ A/cm^2) for the entire surface of the earth is of the order of 1600 A, we must search for the processes which compensate this current. The simplest way of expressing this

is in the form of a current of the same magnitude but opposite direction, flowing away from the earth's surface, namely the so-called counter-current.

For a long time electrical phenomena in the atmosphere were studied mainly under fine-weather conditions without reference to phenomena associated with thunderstorm processes and the question of the causes responsible for the existence of the electric field could not be solved. Many different hypotheses were put forward regarding the sources which provide the earth with the negative charge cancelled by the conduction current. For example, it was thought that various processes occurring in the earth's interior or in the crust might possibly explain the negative charge of the surface. Certain researchers sought these sources in the influx of negative charges from outer space. Numerous suggestions were made in connection with attempts to find a common explanation for the origin of the earth's magnetic field and the electric field of the atmosphere. But all such hypotheses were subsequently rejected and it is now universally acknowledged that the key to the problem lies in the phenomena that take place in the atmosphere itself, and more specifically in the entire assembly of electric currents instrumental in bringing about the exchange of charge between the earth's surface and atmosphere. The first person to point this out was Wilson.

When considering all electric currents with the aim of solving the above problem, one must obviously decide whether the balance of these currents can be reduced to zero for the entire surface of the earth. While most of the surface receives positive charges from the vertical current, the other currents carry both positive and negative charges: as we saw earlier, point discharges and lightning currents carry predominantly negative and moreover considerable charges.

One might expect that at times and places of thunderstorm development the lightning and point-discharge currents would compensate for the negative charge drain due to the conduction current and precipitation over the rest of the surface and that the total current balance would thus be zero. But verification of such considerations requires long systematic observation over a vast area. At individual points observations will naturally give nonzero balances and a positive electricity influx will predominate at some places while a negative one will predominate at others. To illustrate this we present the electricity in coulombs received per square kilometer of surface per year at Cambridge, England:

Conduction current	+60	$C/km^2 \cdot$ annum
Precipitation currents	+20	"
Lightning discharges into earth	-20	"
Point-discharge currents	-100	"
Total	-40	"

These data show that currents from point discharges predominate, an indication confirmed by observations at other places.

The small negative balance observed at middle latitudes must surely be different in the tropics and especially in areas of world thunderstorm foci. Here the number of overland thunderstorms is large and correspondingly the outflow through lightning and point discharges is also large. At the

same time currents from points are excluded over much of the surface— that is, the oceans— and one can suppose that the balance is positive there. Ultimately for the entire surface of the earth the total balance will probably turn out to be zero and the above four processes can then be regarded with complete justification as responsible for maintaining the equilibrium state of the electric currents between the earth's surface and the atmosphere.

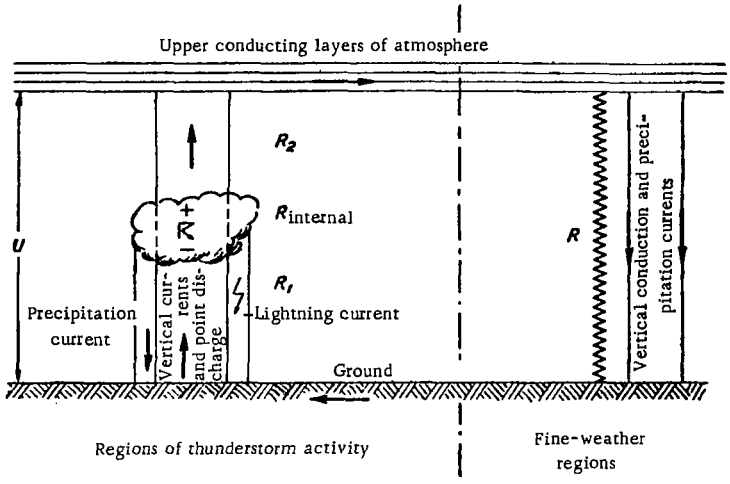


FIGURE 200. Electric currents in the atmosphere

The above statements are represented schematically in Figure 200, which which is a development of the scheme given in Figure 198. The right-hand side of the figure represents all thunderstorm-free regions of the surface; there the currents flowing through the atmosphere, which has the resistance R, are composed of the conduction current and the current of local precipitation. This total current is compensated by currents flowing in the opposite direction in regions of thunderstorm activity (left-hand side). Currents in these regions comprise: 1) the vertical current and leakage currents from points present at the surface; 2) currents of lightning strikes into the ground; 3) local precipitation currents. The ratio between the densities of these currents is different at different instants and different places.

These two branches of the atmospheric electric current system are closed below by currents flowing in the earth's crust and above by currents flowing in the upper conducting layers of the atmosphere, which lie at heights of several tens of kilometers (≈ 60 km) where the conductivity is high enough to ensure rapid dispersal of the charge obtained in any area over the entire earth.

Many researchers have calculated the balances of these currents as averages for the whole earth. So far, however, all such computations remain tentative. Fairly reliable and extensive data are available for the vertical conduction current; for the precipitation current the data, though less reliable, can nevertheless be used as provisional values. The total current density computed for the entire surface can be estimated as roughly

$4 \cdot 10^{-16}$ A/cm^2 for fine-weather areas, which gives a current of the order of 1700–1800 A. To estimate the total current in regions of thunderstorm activity one needs to have access to data concerning the size of these regions, the number of lightning strikes into the ground, the lightning current and current from points, and the conduction and precipitation currents in these regions. Data of this kind are scanty and, in view of the high variability of all these quantities, mean values tend to be unreliable. However, estimates lead to numbers of the same order. Thus if one assumes that there are 100,000 lightning strikes into the ground per hour and every strike carries on the average 20 C of electricity, and if one further recalls that the number of strikes conducting negative charges is roughly four times greater than the number of strikes carrying positive charges, then the density of the lightning current spread over the entire surface of the earth turns out to be of the order of $-0.7 \cdot 10^{-16}$ A/cm^2. If one assumes that the current from points, together with the other currents indicated above, will be 4–5 times greater, as certain observations show, the balance turns out to be zero. Further observations are necessary for greater accuracy but it is perfectly evident from what we said that thunderstorms are the prime generator responsible for maintaining the system of electric currents in the atmosphere and, by the same token, the atmospheric electric field.

Important indirect confirmation of the validity of these considerations is provided by the parallelism between the daily march of the intensity of thunderstorm activity for the whole earth and the unitary variation of the gradient over oceans, on the one hand, (Figure 192), and the daily march of atmospherics, on the other hand.

It should be stressed that these considerations, while establishing the balance between atmospheric currents, entirely fail to shed light on the question of whether the earth and its atmosphere are electrically neutral for outer space; certain studies indicate the possibility of an electrification of the earth and atmosphere, but perhaps this problem will be solved with the help of observations from artificial earth satellites.

Chapter 32

POLAR AURORAS

§ 1. Geographic distribution of auroras and their forms

Polar auroras are a special form of airglow occurring in the upper layers of the atmosphere. They are observed chiefly in polar regions, in the Arctic as well as the Antarctic, and are far less frequent in middle latitudes. Auroral displays have also been reported in exceptional cases nearer the equator, e.g., Istanbul, Athens, Bombay, Aswan and even Singapore (latitude 1° N.), and, in the southern hemisphere, Samoa (13° S.) and Djakarta (6° S.).

Long-term observations have led to the compilation of charts of isochasms (lines of equal auroral frequency). Figure 201 presents such maps for the northern and southern hemisphere (after Vestine). In each hemisphere the isochasms closely follow the circles with centers in the regions of the magnetic poles and the zone of maximal frequency lies at a distance of roughly 20–25° from the pole; away from this zone the frequency decreases both poleward and towards the middle latitudes. On the basis of theoretical considerations and certain observations it may be stated that auroras in the northern and southern hemispheres are closely related and that under favorable conditions displays can be observed simultaneously in both hemispheres.

Auroras have a wide variety of forms, often highly changeable and extremely colorful. The Atlas of Auroral Forms* compiled by Störmer lists several groups falling into two main classes

1) forms without ray structure (quiet forms); these appear as homogeneous arcs or a uniform diffuse surface, and also homogeneous bands; they sometimes pulsate gently;

2) forms with ray structure, as a rule highly mobile; these are represented first and foremost by rays (often colored) diverging from a polar arc, then by rayed bands, and finally by the particularly striking draperies and coronas.

Störmer further distinguishes a third class of flaming auroras; these are extremely mobile forms reminiscent of bursts of flame, moving rapidly in the vertical direction. Photographs of several auroras are given in Figure 202.

The connection between polar auroras and the state of the earth's magnetic field (magnetic storms and other phenomena) was noticed long ago.

* [Superseded by the 'International Auroral Atlas', published for the International Union of Geodesy and Geophysics by the Edinburgh University Press, 1963. The classification given in the latter contains some major changes from Störmer's classification.]

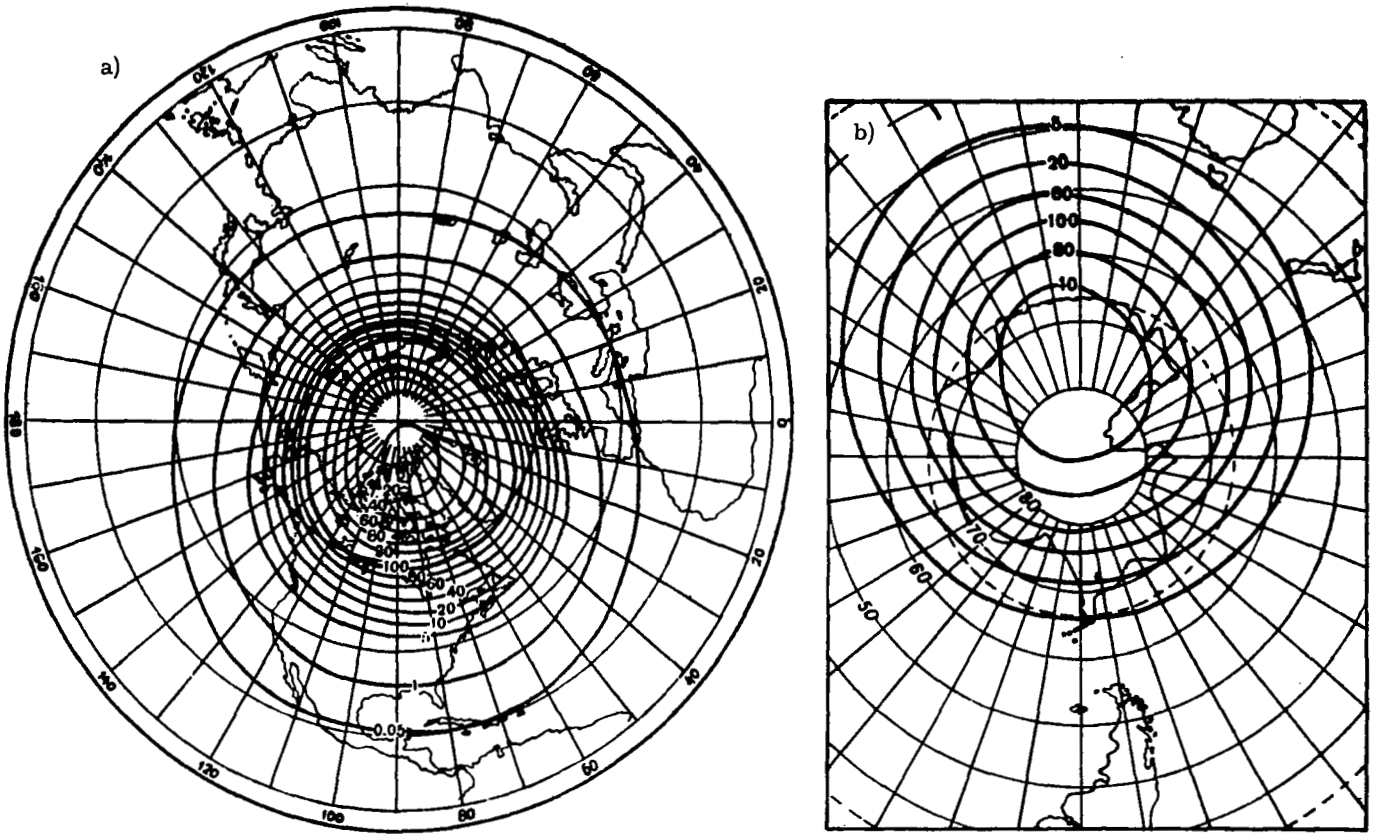


FIGURE 201. Frequency of polar auroras (after Vestine)

a-northern hemisphere; b-southern hemisphere.

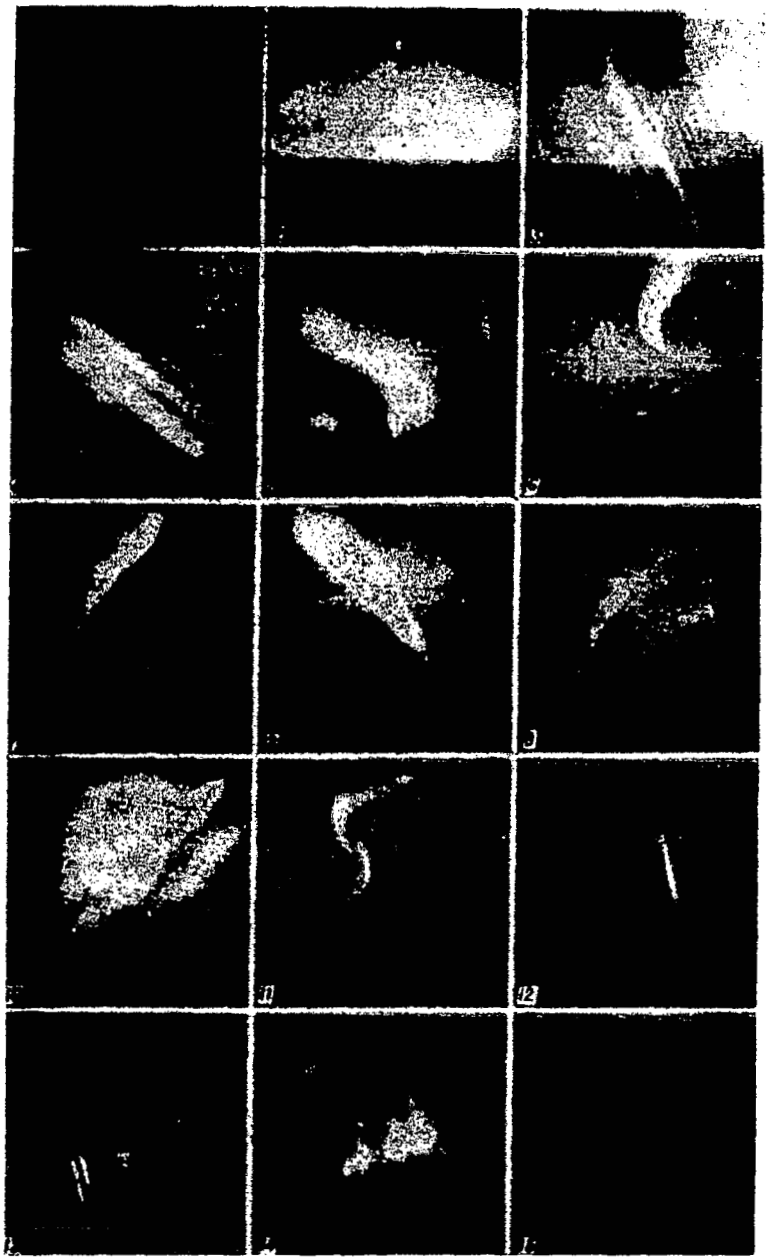


FIGURE 202. Principal auroral forms

1–pulsating surface; 2–quiet arc; 3–split band; 4–split arc; 5–arc; 6, 7, 8–drapery-shaped arc; 9, 10–parallel arcs; 11–drapery-shaped arc; 12–draperies, seen in the tangent plane; 13–draperies with rays; 14–ray-shaped draperies; 15–draperies.

In the last decades a close correlation has also been established with other geophysical phenomena governed by upper-atmosphere processes, including the state of the ionosphere (conditions of radio wave propagation), nightglow and so forth. Furthermore, it has been proved that all these phenomena are closely dependent on the degree of solar activity. We note that the ties linking polar auroras to solar activity and the other phenomena just mentioned are particularly striking in the case of moving ray forms belonging to the second class. One can assume that different auroral forms are produced by different physical processes.

The correlation with sunspot activity is reflected in particular in the fact that auroral displays in years with maximum sunspots are not only more intensive and more frequent but also extend over a much greater distance from the magnetic poles.

§ 2. Height of polar auroras. Their distribution in space. Periodic variations

Base-line observations in conjunction with special methods of photography have established that polar auroras occur over a height range of 80 to 1000–1200 km. The lower edges (irrespective of the form) lie predominantly at heights of 80–160 km (roughly), with a well-defined maximal frequency at 100–106 km. This is clearly shown by Figure 203, compiled after observations in Norway on the basis of 12,330 photographs taken from 1911 to 1944. It is mentioned that the more intensive the display, the lower its bottom level. The height of the upper limit of an aurora varies a great deal depending on the form.

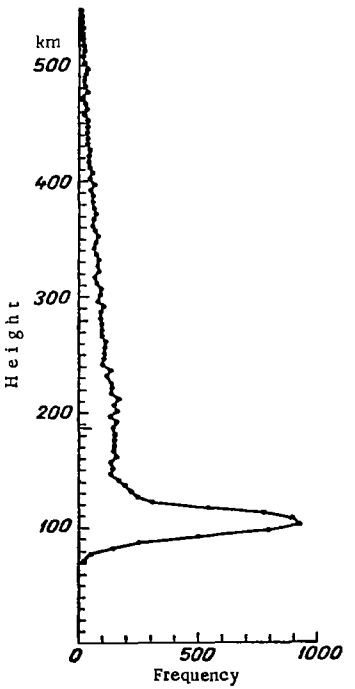


FIGURE 203. Height distribution of polar auroras (after Störmer)

The greatest altitudes are reached by the auroras occasionally observed in parts of the atmosphere still illuminated by the sun's rays. These sunlit auroras differ in many respects from the ordinary displays observed in portions of the atmosphere occurring in the earth's shadow. Their lower edges distinctly follow the boundary separating the lighted and unlighted portions of the atmosphere (200–400 km), while their upper limits extend to considerable altitudes (1000–1100 km). They are observed in areas several hundred kilometers distant from the ordinary auroras, have a grayish-violet or blue coloring and usually appear three hours after and three hours before sunset; as a rule they are accompanied by strong perturbations in the ionosphere and by magnetic storms.

The distribution of auroras in space is simplest to define for such forms as rays. The latter coincide fairly closely with the directions of the lines of force of the constant magnetic field of the earth. Arcs, bands and other forms,

which have considerable horizontal extension are usually so arranged that their horizontal projection is approximately perpendicular to the magnetic meridian. This perpendicular arrangement of the arcs with respect to the horizontal component of the earth's magnetic field is particularly clear-cut at great altitudes.

Statistical processing of the extensive observational material relating to polar auroras has established the existence of a series of periodic fluctuations, including diurnal, annual and eleven-year variations; in addition, the possible existence of variations with a period of 27 and 30 days, and also of one with a period of roughly 55 years, has been noted. However, the fact that for a long time visual techniques have confined observation to relatively clear night skies and, at very high latitudes, to the winter months (especially in the case of feeble displays) has greatly hampered the detection of such variations. Recently the use of radar has made it possible to determine the location of auroral regions in space even in the presence of clouds, thus greatly increasing the possibilities for research into auroral phenomena and study of the upper layers of the atmosphere. One may state that in the daily march of the moving auroral forms the maximum frequency is observed in the evening, roughly 1–2 hours before midnight by local time; quiet forms are reported chiefly in the morning.

In the annual march the maximal frequency is observed near the equinoxes (March and October) in both hemispheres.

§ 3. Color, intensity and spectrum of polar auroras.

When sufficiently bright polar auroras often display rapidly changing colors. Yellowish-green tones usually predominate and are reported in nearly all forms. Bright auroras take on a reddish tone; a faint bluish tone is often observed, and in sunlit portions of auroras even violet tones may be observed. A rapid succession of colors is particularly characteristic of moving forms.

The intensity of auroral glow is highly variable and the illumination produced by auroras at the earth's surface can sometimes exceed the lunar illumination. The intensity usually increases sharply from the lower edge for a vertical distance of ten kilometers, then decreases more or less rapidly with height.

The study of the auroral spectrum is especially important inasmuch as it enables one to draw fairly reliable conclusions regarding the composition of the atmosphere up to great altitudes, the temperatures at these heights and the nature of certain processes taking place there.

The auroral spectrum has much in common with the nightglow spectrum. It is highly complex, containing over 500 lines and bands. The most characteristic and intense lines are those belonging to atomic oxygen: yellow-green lines with a wavelength $\lambda = 5577.34$ Å and red lines with wavelength $\lambda = 6300.3$, 6364 and 6391 Å. As certain authors point out, it is not to be excluded that the auroral spectrum also contains bands belonging to neutral O_2 molecules, such as the band situated between 5528 and 5280 Å.

Lines belonging to the so-called negative system of nitrogen bands emitted by ionized molecules of nitrogen (N_2^+) have a pronounced intensity in the auroral spectrum. The most important of these are lines

in the violet and ultraviolet regions of the spectrum with wavelengths of 4708, 4279 and 3914 Å. Furthermore, there are systems of bands emitted by neutral nitrogen molecules. These comprise the first positive system, observed in the red and infrared regions of the spectrum with principal bands of 5991 and 5867 Å, and also the fainter second positive system, dominant in the ultraviolet region with principal bands near 3997 and 4059 Å. Further, neutral nitrogen atoms emit a weak system, the so-called Vegard-Kaplan bands, in the violet and ultraviolet regions of the auroral spectrum.

The auroral spectrum, like that of nightglow, displays the yellow line of neutral sodium atoms (5890 and 5896 Å), as well as bands of the hydroxyl molecule. Helium lines are not found in auroral spectra. The presence of hydrogen lines (H_α, H_β and H_γ), on the other hand, has been reliably established in a number of cases; their Doppler broadening indicates that the hydrogen atoms present in the atmosphere are in very vigorous motion, with speeds which, according to certain observations, reach up to about 3500 km/sec.

It should be mentioned that the spectra of individual auroras vary; in many cases certain lines are altogether absent while others are very prominent.

A series of correlations has been noted between the character of the spectrum, on the one hand, and the form and height of the aurora on the other; this is partly reflected in color changes as well. Without going into details we note that in the case of auroras appearing in the shaded and the sunlit portion of the atmosphere the intensity of the green line ($\lambda = 5577$ Å) in the sunlit portion is considerably weaker than that of the red line of oxygen ($\lambda = 6300$ Å) and of the lines of nitrogen (4278 and 3914 Å).

Numerous studies have been made of the complex character of the auroral spectrum and its peculiarities, and attempts have been made to explain them. However, many unsolved problems remain. The above-mentioned similarity between the spectra of auroras and nightglow is perfectly understandable in view of the fact that both phenomena are determined by processes occurring in the same region of the atmosphere. But at the same time important differences between them indicate that the mechanisms which activate them are different. All this calls for further detailed investigation.

§ 4. Theory of polar auroras

It was already mentioned by Lomonosov that polar auroras resulted from the action of electrical forces and were similar to the phenomenon of electrical discharge. Paulsen suggested more positively that auroras were generated by a flux of cathode rays emanating from the sun and penetrating into the upper atmospheric layers. At the end of the last century (1896) Birkeland was able to demonstrate this possibility experimentally. In Birkeland's experiments a sphere was placed inside a large cathode tube (volume about 1 m^3); this sphere, consisting of an electromagnet covered with a phosphorescent shell, was named "Terrella" by Birkeland. Upon exposure of the magnetized "Terrella" to cathode rays for different positions of the cathode a number of phenomena analogous to the auroras

observed in nature could be discerned. Subsequently (1934) these experiments were repeated by Brüche using more advanced techniques.

From Birkeland's experiments it followed that under the influence of the earth's magnetic field the stream of corpuscles emanating from the sun may under certain circumstances bend around the earth and concentrate near its poles, leading to the glowing of rarefied gases in the upper atmospheric layers. It was necessary to determine theoretically the conditions under which the corpuscular rays from the sun could reach the earth and what their trajectory would be. The solution of this problem was first given by Störmer. In view of the complexity of the problem, Störmer initially introduced the following simplifications:

FIGURE 204. Trajectories of particles

1) the velocity of the corpuscles is so large that in the time they take to move from the sun to the earth the relative position of sun and earth does not change;

2) the particles are not subjected to any force other than the earth's magnetic field; the influence of the sun's magnetic field is not considered;

3) the particles move in thin streams which do not influence each other;

4) the earth's magnetic field can be treated as the field of a uniformly magnetized sphere or a dipole placed at the origin of coordinates.

Numerical calculations enabled Störmer to construct spatial models for a large number of particle trajectories, analysis of which revealed that of the corpuscles emanating from the sun only a very few having a completely definite, so-called favorable, direction would reach the origin of coordinates (provided they were not absorbed). In general particles will move in the magnetic field along a spiral curling about the magnetic line of force as shown in Figure 204. Thus all corpuscles moving out along directions even very slightly different from the "favorable" one will describe helical trajectories around the magnetic line of force and after penetrating a certain distance into the atmosphere will turn and go back, provided they have not been absorbed in the atmosphere.

On the basis of his calculations Störmer was able to explain the general picture and also various details of the phenomenon, give certain quantitative characteristics, and elucidate the conditions associated with the appearance of certain basic auroral forms (arcs, rays, draperies).

Naturally, inferences drawn from calculations of the motion of a single charged particle and indeed the entire theory based on the above simplifications were not always able to furnish satisfactory replies to certain questions. Thus further refinement was needed. The first major difficulty arose in the theoretical explanation of the appearance of auroras at large angular distances from the earth's magnetic poles (at low latitudes). According to the theory, the angular distance θ to which corpuscles can penetrate into the atmosphere is given by

$$\sin\theta = 2a\sqrt{\frac{mv}{Me}},$$

where m, v and e are respectively the mass, velocity and charge of the particle, M is the earth's magnetic moment and a is the distance from the center of the earth to the height of the lower edge of the aurora.

Calculations with this formula give values of about 2–3° for θ if the active particles are electrons moving at speeds of the order of thousands of kilometers per second. Equally small values are obtained for other charged particles present in the corpuscular stream (protons H^+, calcium ions Ca^+ and so on), which move at the same speeds. Yet observations indicate appreciably higher values for the angular distance of the zone of maximum auroral frequency (about 20–25°), and auroras are often observed, as we mentioned earlier, at even greater distances from the magnetic poles. However, it was possible to obviate this contradiction between theory and observed data by showing that some of the corpuscles emanating from the sun form a ring current around the earth in the equatorial plane at a distance of 6–7 terrestrial radii from its surface. The action of this ring current draws the corpuscles away from the polar and into the middle latitudes; moreover the stronger this current the greater the effect.

Another defect of the theory is that it fails to explain how particles of like charge, forming, as assumed by this theory, a narrow beam, could reach the earth without being scattered at all perceptibly by electrostatic repulsion. It was therefore suggested by Chapman and Ferraro that the corpuscular stream ejected from the active regions of the sun and responsible for auroras was as a whole neutral and consisted of electrons and positively ionized atoms. Such a stream would not be scattered by electrostatic repulsion. As the stream approached the earth at a mean speed of the order of 1000–2000 km/sec, at a distance of several earth radii the positively and negatively charged particles would be deflected in opposite directions — the former westward and the latter eastward — by the earth's magnetic field, and would be decelerated. As a result the earth would lie, as it were, in the hollow of this stream, with opposite charges concentrated on opposite walls. This would induce an electric field which would cause the charges to travel from one wall to the other, enclosing the hollow and forming a circular current flowing westwards in the equatorial plane of the earth. This theory gives a satisfactory explanation for the formation of the equatorial ring current and thus for the broadening of the auroral zone but scarcely touches the fundamental problem of the origin of the auroras themselves; the latter, according to this theory, are produced not by direct collisions between solar corpuscles and the molecules and atoms of the upper atmosphere but as a consequence of electrical discharges between the walls of the hollow in the corpuscular stream and the ionosphere. Further extension of the theory suggests that even before the equatorial ring current is closed some of the corpuscles constituting the neutral stream may penetrate into the polar regions, producing polar auroras there.

A number of considerations regarding the possibility of deep penetration of the earth's atmosphere by corpuscles has been put forward recently. In 1951, for example, I. S. Shklovskii advanced a hypothesis based on the mechanism of the so-called charge exchange of solar protons upon collision with atoms and molecules of the terrestrial atmosphere; as a result of this process they would move as neutral hydrogen atoms along part of the way. A great deal of attention is paid also to the question of the formation of the corpuscular stream and its propagation in outer space. According to the theory of Bennett and Hulburt (1954), the corpuscular stream,

remaining near-neutral at all times, is propagated in the form of a narrow jet owing to the so-called magnetic self-focussing. As we are unable to examine these and other theories in any detail, we confine ourselves to a brief note concerning the so-called "ultraviolet" theory. According to this theory, the ionized particles responsible for polar auroras are formed in the atmosphere itself as a result of ionization by ultraviolet rays of the sun during violent ultraviolet flares. The ionization of particles takes place at very great heights (tens of thousands of kilometers), which, as pointed out in Chapter 4, can be reached by molecules and atoms of atmospheric gas having very long mean free paths in view of the high rarefaction of the upper layers. On returning towards the surface the ionized particles are deflected by the magnetic field towards the polar regions where they would be able to produce polar auroras. This theory explains certain experimental data fairly successfully but is open to a great many serious objections. Now that the existence of solar corpuscular streams has been proved the "ultraviolet" theory has been abandoned by most investigators.

In conclusion one can say that while no complete theory of polar auroras is yet in existence it is likely that the final theory will be based on the Birkeland-Störmer theory with due allowance for those features which will eventually be established for the electrically "quasi-neutral" streams of corpuscles emitted by the sun.

Part Eight
ATMOSPHERIC ACOUSTICS

The study of the propagation of acoustic vibrations in the atmosphere under natural conditions constitutes the subject matter of the branch of meteorology known as atmospheric acoustics.

In speaking of acoustic vibrations we shall henceforth refer to vibrational motions of small amplitude usually described as sound vibrations and perceptible to the human ear (frequency of 40 to 20,000 hertz).

The question of the propagation of sound waves in the atmosphere is exceedingly complex, since the atmosphere is not a homogeneous acoustic medium and is constantly changing properties; moreover it is not immobile and therefore all phenomena must be considered in a moving medium.

The study of the propagation of sound in the atmosphere, aside from its great practical importance for problems of acoustic signalling, the location of sound sources and so forth, is also important for the study of the atmosphere itself insofar as it sheds light on the structure of the atmosphere at great altitudes and on processes occurring in the lower layers

Chapter 33

SPEED OF SOUND IN THE ATMOSPHERE

§ 1. Speed of sound in an immobile homogeneous medium

The propagation of sound in gases can be treated as a particular case of the propagation of elastic vibrations in a medium in which there is no rigidity and in which the forces of buoyancy in the undeformed state are not zero but reduce to a uniform hydrostatic pressure p_0. Only longitudinal waves will be propagated in such a medium. For the elastic stresses inside it one need consider only the pressure variations arising on elastic deformation of the volume, i.e., the quantity $p-p_0=\delta p$. Assume that the excess pressure is positive for compression and negative for expansion and that, like the changes in the other parameters of the medium, it is very small for the sound vibrations under consideration. Then, in accordance with the theory of sound propagation in a homogeneous immobile medium in the absence of body forces, we write the wave equation

$$\frac{\partial^2\theta}{\partial t^2}=c^2\nabla^2\theta, \tag{1}$$

where θ is the volume deformation and ∇^2 the Laplacian operator:

$$\theta\approx\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}.$$

Here u, v and w are the components of the particle displacement along the coordinate axes and c is the speed of propagation of vibrations, given by

$$c=\sqrt{\frac{\delta p}{\delta\rho}}, \tag{2}$$

where ρ is the density of the medium.

Replacing the quotient of δp and $\delta\rho$ in the above (in view of the smallness

of these quantities) by the derivative $\frac{dp}{d\rho}$, we have

$$c = \sqrt{\frac{dp}{d\rho}}. \tag{3}$$

Making use of the equation of polytropy $pv^k = \frac{p}{\rho^k} = a = \text{const}$, from which $\frac{dp}{d\rho} = ak\rho^{k-1} = k\frac{p}{\rho}$, we obtain in place of (3) the expression

$$c = \sqrt{k\frac{p}{\rho}}. \tag{4}$$

One can imagine two cases of the propagation of sound vibrations: 1) the vibrations are so slow that the entire process takes place isothermally and the temperature succeeds in equalizing itself between the zones of compression and rarefaction in the wave; 2) the vibrations are so rapid (period < 1 sec) that the process takes place adiabatically and changes in air temperature take place due to compression and rarefaction. In the former case, as is known, $k = 1$ and therefore the speed of sound is given by

$$c = \sqrt{\frac{p}{\rho}}. \tag{5}$$

This is Newton's formula, from which at normal pressure ($p = 760$ mm Hg) and temperature ($T = 273°$) we obtain the numerical value $c = 280$ m/sec; this, however, is considerably smaller than the observed value.

In the second case, where the process is adiabatic, the polytropic exponent is $k = \varkappa = \frac{c_p}{c_v}$, where c_p and c_v are the heat capacity at constant pressure and constant volume respectively. The speed of sound is then given by

$$c = \sqrt{\varkappa\frac{p}{\rho}}. \tag{6}$$

This is Laplace's formula, from which we have $c = 331.8$ m/sec for dry air ($\varkappa = 1.405$) at normal temperature and pressure; this coincides with observational results.

Let us write formula (6) in a slightly different form, making use of the equation of state $p = \rho\frac{R^*T}{\mu}$; then

$$c = \sqrt{\varkappa\frac{R^*T}{\mu}}, \tag{7}$$

which shows that the speed of sound depends only on the nature of the gas ($\varkappa$ and μ) and its temperature.

§ 2. Dependence of the speed of sound in the atmosphere on temperature and humidity

When applying formula (6) to the actual atmosphere one must allow for the influence of atmospheric humidity and temperature, on which both $\varkappa$ and ρ depend. The temperature dependence of $\varkappa$ is small and can be

disregarded. The humidity dependence of $\varkappa$ is expressed by the relation

$$\varkappa = \varkappa_1 \frac{1 + 0.280 \frac{e}{p}}{1 + 0.393 \frac{e}{p}}, \tag{8}$$

where $\varkappa_1$ is the value of $\varkappa$ for dry air ($\varkappa_1 = 1.405$).

The density of air is related to the humidity and temperature by the well-known expression

$$\rho = \rho_0 \frac{273\left[1 - 0.378 \frac{e}{p}\right]}{T}, \tag{9}$$

where ρ_0 is the density of dry air at $T = 273°$ at the pressure p and e is the pressure of water vapor.

Introducing (8) and (9) into (6), we obtain

$$c = \sqrt{\frac{p}{\rho_0} \varkappa_1 \frac{T\left[1 + 0.280 \frac{e}{p}\right]}{273\left[1 - 0.378 \frac{e}{p}\right]\left[1 + 0.393 \frac{e}{p}\right]}}. \tag{10}$$

From the above for dry air ($e = 0$) at $T = 273°$ we have, as before,

$$c = c_{0,\ \mathrm{dry}} = \sqrt{\varkappa_1 \frac{p}{\rho_0}} = 331.8 \text{ m/sec},$$

for dry air at any temperature

$$c = c_{T,\ \mathrm{dry}} = \sqrt{\varkappa_1 \frac{p}{\rho_0} \frac{T}{273}} = c_{0,\ \mathrm{dry}} \sqrt{\frac{T}{273}} = 20.1 \sqrt{T}. \tag{11}$$

Finally, for moist air at temperature T and humidity e we have formula (10), which upon simplification can be written in the form

$$c = \sqrt{\varkappa_1 \frac{p}{\rho_0} \frac{T}{273}\left[1 + 0.265 \frac{e}{p} + 0.144 \left(\frac{e}{p}\right)^2\right]}, \tag{12}$$

or, as it is usually taken for practical calculations,

$$c = \sqrt{\varkappa_1 \frac{p}{\rho_0} \frac{T}{273}\left(1 + 0.275 \frac{e}{p}\right)} = c_{0,\ \mathrm{dry}} \sqrt{\frac{T\left(1 + 0.275 \frac{e}{p}\right)}{273}}. \tag{12'}$$

If we introduce the so-called virtual acoustic temperature by the relation

$$T_{\mathrm{v.a.}} = T \sqrt{1 + 0.275 \frac{e}{p}},$$

we then obtain, by analogy with (11),

$$c = c_{0,\ \mathrm{dry}} \sqrt{\frac{T_{\mathrm{v.a.}}}{273}} = 20.1 \sqrt{T_{\mathrm{v.a.}}} \text{ m/sec}. \tag{12''}$$

From the above argument it is evident that:

1) the speed of sound does not depend explicitly on the density of air;

2) in dry air the speed of sound is proportional to the square root of the absolute air temperature and one can find by calculation that for a temperature rise of 1° the speed increases roughly by 0.6 m/sec;

3) the speed of sound in moist air depends on the ratio $\frac{e}{p}$ rather than on the individual quantities; this dependence is not large and, as calculations show, for $e < 7$ mm disregarding it produces errors not exceeding 0.5 m/sec.

In view of this for approximate calculations use is often made of the formula

$$c = 331 + 0.6t + 0.07e, \tag{13}$$

which for temperatures between −20 and + 30° gives an error of no more than +0.5 m/sec.

Since in the atmosphere the temperature and humidity change with height, the speed of sound also changes with height; obviously, it will decrease in those atmospheric layers in which the temperature drops with height.

§ 3. Speed of sound in the presence of wind

When using the above relations to calculate the speed of sound in the atmosphere one must take into account that the atmosphere is not an immobile medium. Let us see how the wind influences the rate of propagation of sound. We shall confine ourselves to the case of a medium moving at constant speed. If the observer is moving together with the medium at the same speed he will see the phenomenon as it would be in an immobile medium and the initial equation would remain valid. However, if the observer is standing on the ground and is not moving with the air, he will see a different picture.

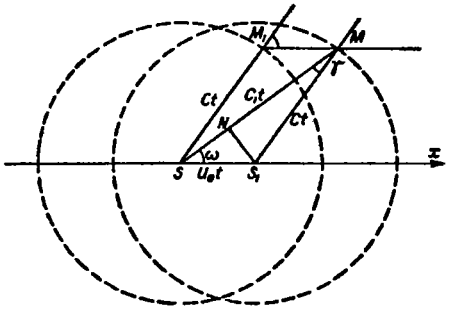


FIGURE 205

To show this, consider the propagation in the horizontal plane of a sound wave generated by a source situated at the point S (Figure 205). If the atmosphere were immobile then obviously the wave, propagating in a uniform medium at the same speed c in all directions, would simultaneously reach points located on a circle of radius $SM_1 = ct$ after the time t. In the presence of wind when the medium is moving along the x-axis at a speed u_0 all vibrating particles also move in this direction; the point M_1, for instance, will be shifted to the point M by the distance $MM_1 = u_0t$. An observer standing at the point M will perceive the sound as if it were coming from the point S_1 but, referring to the location of the source of the sound S

he will estimate its velocity from the time spent on travelling the path $SM = c_1 t$ and, furthermore, will notice a deflection by the angle γ of the direction to the source S from the direction along which he perceived the sound.

Denoting by ω the angle between the wind direction (x-axis) and the direction to the location of the sound source MS, one can write

$$SM = NM + SM = S_1M \cos\gamma + SS_1 \cos\omega,$$
$$c_1 t = ct \cos\gamma + u_0 t \cos\omega, \tag{14}$$

or

$$c_1 = c\cos\gamma + u_0 \cos\omega \tag{15}$$

However, for any value of ω and even for large wind speeds the angle γ is so small that one can assume with sufficient accuracy that $\cos\gamma = 1$.

Indeed, from the triangle SS_1M we have

$$\sin\gamma = \frac{SS_1}{S_1M}\sin\omega = \frac{u_0}{c}\sin\omega,$$

which even for $\omega = 90°$ gives the maximum value of $\sin\gamma$ as the ratio of the wind speed to the speed of sound, which is always small. Therefore the angle ω can be set equal to the angle formed by the wind direction at the point of observation M with the observed direction of sound arrival S_1M. In view of this one can write

$$c_1 = c + u_0 \cos\omega. \tag{16}$$

Thus in the presence of wind the observed speed of sound c_1 will depend on the direction of propagation of the sound with respect to the wind direction; it will be maximum in the direction of the wind and minimum in the opposite direction.

Chapter 34

PROPAGATION OF SOUND IN THE ATMOSPHERE

§ 1. Path of a sound ray in the atmosphere

Owing to the inhomogeneity of the atmosphere in both the vertical and the horizontal directions, the speed of sound, which is dependent on temperature, humidity and wind, will be different in different parts of the atmosphere and therefore the wavefront will be changing. When considering the propagation of sound in the atmosphere it is more convenient to adopt the viewpoint of geometrical optics. We shall proceed to elucidate this, confining ourselves to the simplest cast in which the speed of sound is a function only of the height above ground level.

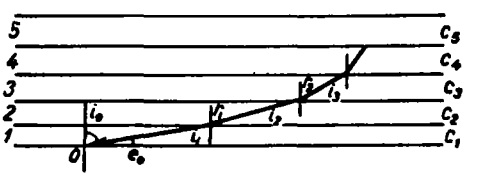


FIGURE 206

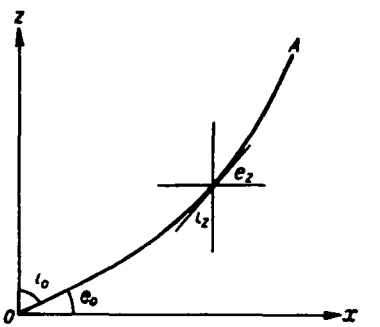


FIGURE 207

Disregarding the curvature of the earth's surface, let us derive the equation of the path of a sound ray in such a smoothly inhomogeneous medium. To do this we divide the atmosphere into a series of layers (Figure 206) in each of which we consider the speed of sound to be constant. Obviously, on moving from one layer to the next the sound ray will be refracted. Denoting the angles of incidence by $i_1, i_2, i_3 \ldots$ and the angles of refraction by $r_1, r_2, r_3 \ldots$, we write

$$\frac{\sin i_1}{\sin r_1} = \frac{c_1}{c_2}, \quad \frac{\sin i_2}{\sin r_2} = \frac{c_2}{c_3}, \text{ etc.}$$

and since $r_1 = i_2$, $r_2 = i_3$ and so on, we have

$$\frac{\sin i_1}{c_1} = \frac{\sin i_2}{c_2} = \ldots = \frac{\sin i_k}{c_k} = \text{const} = K = \frac{\sin i}{c}. \qquad (1)$$

Introducing the angle $e=90°-i$, which is called the angle of emergence, we obtain

$$\frac{\sin i_z}{c_z}=\frac{\cos e_z}{c_z}=\frac{\sin i_0}{c_0}=\frac{\cos e_0}{c_0}=K, \tag{1'}$$

where quantities referring to the earth's surface are indicated by the subscript 0.

From (1') it is shown that, since for the vertical temperature lapse usually observed $c_z<c_0$, $\cos e_z<\cos e_0$ and consequently $e_z>e_0$, i. e., the ray is deflected upwards when propagating upwards.

Consider the equation of the path of a ray, written in differential form (Figure 207) as follows:

$$\operatorname{tg} i=\frac{dx}{dz}=\frac{\cos e_z}{\cos i_z}=\frac{Kc_z}{\sqrt{1-K^2c_z^2}} \tag{2}$$

Integrating the above along the path of the ray from the point O to the point A, we obtain

$$x=\int_0^A \frac{Kc_z\,dz}{\sqrt{1-K^2c_z^2}}. \tag{3}$$

Let us consider the most frequent case, in which it can be assumed that the temperature in the atmosphere varies according to the linear law $T_z=T_0-\gamma z$. In this case equation (2) becomes

$$\frac{dx}{dz}=\operatorname{tg} i=\frac{\sin i}{\cos i}=\frac{1}{\sqrt{\frac{1}{\sin^2 i}-1}},$$

and since on the basis of (1')

$$\sin i=\frac{c_z}{c_0}\sin i_0=\sqrt{\frac{T}{T_0}\sin i_0}=\sqrt{1-\frac{\gamma z}{T_0}}\sin i_0,$$

we have

$$\frac{dx}{dz}=\frac{1}{\sqrt{\frac{1}{\left(1-\frac{\gamma z}{T_0}\right)\sin^2 i_0}-1}}. \tag{4}$$

One can easily show by integrating the above that the path of the ray is a cycloid. However, one can assume approximately that $\frac{1}{1-\frac{\gamma z}{T_0}}=1+\frac{\gamma z}{T_0}$; then the equation of the path (4) becomes simpler:

$$\frac{dx}{dz}=\frac{1}{\sqrt{\frac{\gamma z}{T_0}}}. \tag{5}$$

Integrating the above we obtain

$$x=2\sqrt{\frac{T_0 z}{\gamma}}, \tag{6}$$

i. e., the path of the ray can be represented approximately as a parabola.

The expressions obtained show that the nature of the path depends on the sign and magnitude of the vertical temperature gradient. Indeed, according

to (4), at the point where total internal reflection of sound takes place, i. e., where $i = 90°$ and $\operatorname{tg} i = \infty$, we have

$$\left(1 - \frac{\gamma z}{T_0}\right) \sin^2 i_0 = 1, \tag{7}$$

whence we obtain the height of this point, which is given by

$$z = z' = -\frac{T_0}{\gamma} \cos^2 i_0. \tag{8}$$

Consequently, for $\gamma > 0$ (temperature decreasing with height) $z' < 0$, i.e., the ray is convex towards the earth's surface; for $\gamma < 0$, i. e., in the presence of inversion, the ray reverses its path; only for $i_0 = 90°$ (horizontal ray) is $z' = 0$.

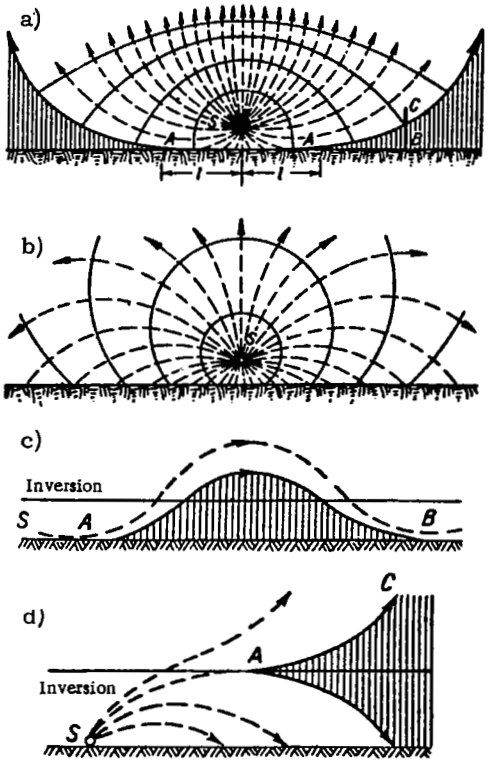


FIGURE 208. Bending of sound rays

a–temperature decreases with height; b–temperature increases with height; c–inversion in free atmosphere; d–surface inversion

Let us consider the paths of the sound ray in the absence of wind for different temperature distributions in the atmosphere. We will confine our attention to a series of typical examples, postulating that the sound source S lies at a certain height h above the earth's surface. If the temperature

drops with height ($\gamma>0$) and consequently the speed of sound decreases with height, the rays will be convex towards the earth's surface (Figure 208a). Rays emerging from the point S will reach the surface only within a circle of radius l, equal to the distance from the point O (projection of the point S on the surface) to the point A, at which a certain limiting ray osculates with the surface. This distance can be found from (6): $l=x=2\sqrt{\frac{T_0 h}{\gamma}}$; no ray emerging from the point S will reach beyond it.

At distances greater than l one will observe a zone of silence or sound shadow, indicated on the figures by shading. If we were to stand at the point B in the sound shadow zone and then move upwards, at a certain height we would again enter the zone of audibility (point C).

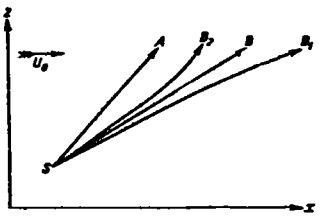


FIGURE 209

Figure 208b shows the ray paths for $\gamma<0$. In this case there is no silent zone since all rays emerging from the point S are concave towards the surface. The other figures show the ray paths when an inversion occurs at a certain height (Figure 208c) and in the presence of a surface inversion (Figure 208d).

Let us now consider the influence of wind on the path of the ray. The simplest case is when the wind direction does not change vertically and the path of a ray is a plane curve. Looking at the propagation of sound in the vertical plane coinciding with the wind direction, one sees at once that if the temperature did not change with height the path of the ray would be a straight line SA (Figure 209) in the absence of wind; in the presence of a constant wind with height, the path would be given by the straight line SB.

If the wind speed increases with height the path will assume the form of the curve SB_1, concave towards the wind direction; if the wind speed decreases vertically the path will be concave in the opposite direction, i.e., the ray will always be deflected in the direction of the wind gradient. This general statement helps explain the nature of the wind's influence from the qualitative standpoint.

When the temperature decreases vertically in the atmosphere the wind's influence produces the changes represented in Figure 210. If the speed of the wind, shown blowing to the right in the figure, increases with altitude, all rays experience an increasingly strong deflection to the right with height; as a result the sound shadow draws nearer to the point O on the left-hand side, while on the right it moves away: when the wind speed decreases with height the deflection pattern is reversed.

Often the wind change has a greater influence on the direction of the ray than the temperature lapse: the paths then assume the forms illustrated by Figure 210a and 210b (wind rises and falls with height, respectively). Since on the average the wind speed rises with height more often than not, sound is usually propagated over a greater distance by the wind. When the vertical distribution of the wind speed is more complex the paths, of course, assume a more complex form. For illustration Figure 210c shows the paths when the wind speed increases to a certain height, and subsequently decreases.

The above discussion was confined to the sound distribution in the vertical plane coincident with the wind direction. If one were to take a plane at an angle to the wind vector the picture would change to some extent though

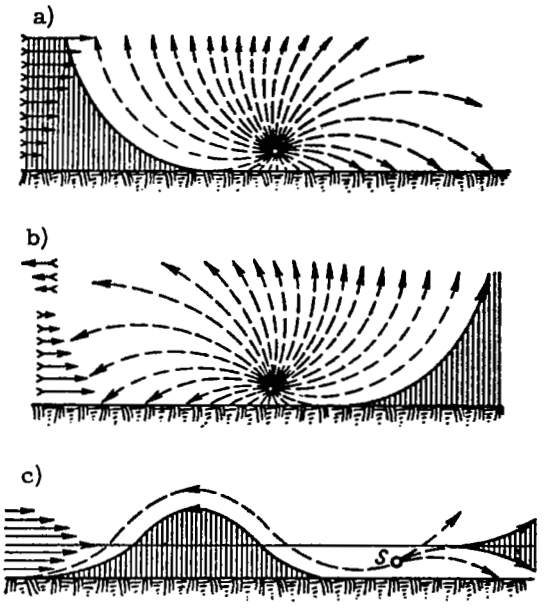


FIGURE 210. Bending of sound rays

a–wind increases with height; b–wind drops with height; c–wind increases with height, and subsequently drops

the general features would remain the same. By simultaneously considering the influence of wind and temperature one can reconstruct the distribution of the silent zones and zones of audibility for the earth's surface for a source situated at a certain height, assuming, of course, the absence of diffuse reflection and other factors which slightly modify the picture.

§ 2. Reflection, refraction and absorption of sound in the atmosphere

If a sound wave strikes any surface, in particular the surface of discontinuity between two atmospheric layers having different properties, it may experience reflection and refraction. In this case laws analogous to the laws of geometrical optics will hold true; according to these laws: 1) incident, reflected and refracted rays lie in the same plane perpendicular to the plane of discontinuity; 2) the angle of incidence i is equal to the angle of reflection i_1; and 3) $\sin i : \sin r = c_1 : c_2$, where i is the angle of incidence, r the angle of refraction, and c_1 and c_2 are the speed of sound in the first and second layer, respectively.

Since the speed of sound depends very slightly on wavelength (with the exception of ultrasound), dispersion is not observed in refraction. We note also that if the sound moves from a medium in which its speed is lower to one in which its speed is higher (to a medium which is acoustically less dense), it will move away from the normal. The theory of reflection and refraction of sound was worked out in detail by Lord Rayleigh. If R, D, and E are the amplitudes of the reflected, refracted and incident waves, respectively, according to the theory we obtain the following relations:

$$\frac{R}{E} = \frac{\frac{\rho'}{\rho} - \frac{\operatorname{ctg} r}{\operatorname{ctg} i}}{\frac{\rho'}{\rho} - \frac{\operatorname{ctg} r}{\operatorname{ctg} i}} \quad \text{and} \quad \frac{D}{E} = \frac{2}{\frac{\rho'}{\rho} - \frac{\operatorname{ctg} r}{\operatorname{ctg} i}}, \tag{9}$$

where ρ and ρ' are the densities of the first and second medium.

These formulas can be used to calculate the reflection of sound both from the surfaces of various objects on the ground (rocks, buildings, etc.) and from surfaces separating different regions in the atmosphere. The latter may be a cloud surface, the surface of a rainfall zone, the surface of discontinuity between two thermally different layers, etc. Such reflections can explain such phenomena as echo, thunder rolls and others.

The propagation of sound vibrations in the atmosphere is accompanied of course by the scattering of energy, and the intensity of the sound decreases with distance from the source. This occurs firstly because the energy emitted by the source is being distributed over increasingly large spherical surfaces, as a result of which the intensity decreases inversely as the square of the distance r from the source, and secondly because the medium itself absorbs energy from the ray. The general law of attenuation of energy for a spherical wave can be written as

$$I = \frac{I_0}{r^2} e^{-2\alpha r} = \frac{I_0}{r^2} p^r, \tag{10}$$

where α is the damping coefficient for the amplitude of the sound wave and $p = e^{-2\alpha}$ can be called the acoustic transmission coefficient.

In the case of a plane wave equation (10) becomes

$$I = I_0 e^{-2\alpha r} = I_0 p^r. \tag{11}$$

In the classical theory evolved by Lord Rayleigh, Stokes and others, the loss of energy is due to the influence of viscosity, heat conduction and thermal emission in the medium. However, when applied to the question of sound propagation in the atmosphere this theory leads to results which are very different from the observed data. The observed values of the damping coefficient are about $3.5 \cdot 10^{-5}$ cm^{-1} and are several times larger than the theoretical ones, which accounts for the very strong attenuation of sound with distance.

The reason for the divergence between theoretical and experimental data can be sought in the structural inhomogeneity of the atmosphere — the presence inside it of suspended foreign particles, solid as well as liquid, and also the presence of inhomogeneities of density, temperature, wind, etc. due to its turbulent state. The presence of foreign particles with dimensions smaller than the wavelength should obviously lead to scattering of the sound waves on them.

The influence of turbulence on the attenuation of sound can also be interpreted as a result of the scattering of the sound energy by eddies and predominantly by those eddies with dimensions commensurate with the wavelength.

Theoretical studies carried out with allowance for all these factors lead to results that are fairly well supported by experimental data. Indeed, calculations give values of the damping coefficient of the same order (10^{-5} cm^{-1}) as the observations.

Thus one can say that the principal factor responsible for the attenuation of sound in the real atmosphere is its inhomogeneity. The determination of quantitative correlations between the various parameters characterizing the atmosphere's inhomogeneity and the values of the damping coefficient is of considerable interest. In particular, by using acoustic methods one can determine the size of turbulent formations.

The following fundamental conclusion may be drawn from the above discussion: the more stable and homogeneous the atmosphere, the smaller the loss of sound energy inside it. It is thus not surprising that sound attenuation should be smaller in winter and at night than in summer and in daytime; similarly, it is less over water and snow surfaces than over land, i.e., the different conditions of audibility in the atmosphere are due to many meteorological factors.

§ 3. Utilization of observations on sound propagation for research into the upper atmosphere

Observation of the propagation of sound waves produced in explosions, volcanic eruptions, artillery fire and other similar powerful sources usually yields the following picture. Around the source of sound there is a

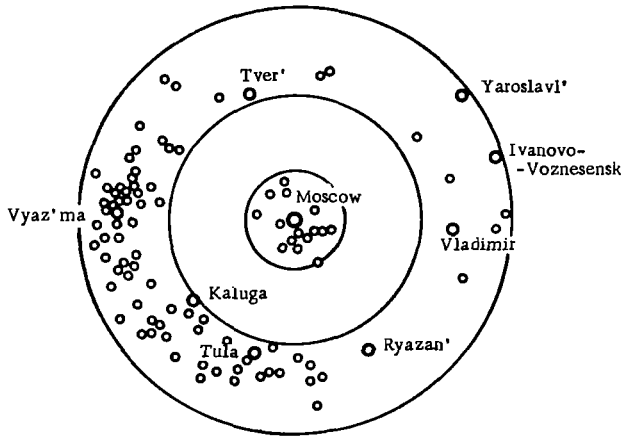


FIGURE 211. Audibility zones during explosion at Moscow in December 1920 (after V.I. Vitkevich)

more or less extensive zone of audibility known as the normal or inner audibility zone, beyond which begins the silent zone. But at even greater

distances the silent zone ends and a new audibility zone, described as the outer or abnormal zone, begins. The distribution of audibility zones during an explosion at Moscow in 1920 is illustrated by Figure 211. This example corresponds closely to the description just given. However, the shape of the boundaries can often depart very strongly from the roughly concentric circles centered about the location of the sound source. The dimensions of these zones also vary. Often the zones are displaced in some direction and are discontinuous. Frequently the second, outer zone of audibility is again followed by a zone of silence beyond which the sound may become audible once more.

It is most important to explain the formation of the outer audibility zone. Obviously, this zone can form only when the sound waves turn back towards the ground on reaching some considerable height, striking the surface in the outer audibility zone. By studying these phenomena one can draw certain inferences concerning the state of the atmosphere at the heights at which reversal of the rays takes place. The problem is not a simple one, of course, especially if one considers the variability and diversity of the factors affecting the propagation of sound. Without going into details, we will attempt to give an idea of the theoretical foundations of the acoustic method of atmosphere research. Let us assume that the atmosphere is immobile and that its physical properties vary only with height. In this case, as we saw, the equation of the path of a ray can be written as

$$\frac{\cos e_z}{c_z} = \frac{\cos e_0}{c_0} = K.$$

If the ray turns towards the surface at some height, then, obviously, at the highest point it will assume a horizontal direction, i.e., here the angle $e_z = 0$; therefore at the highest point, where the speed of sound $c_z = \bar{c}_z$, we will have

$$\bar{c}_z = \frac{c_0}{\cos e_0}. \qquad (12)$$

This shows that if one knows c_0 for rays directed at various angles e_0 to the earth's surface one can find $\bar{c}_z$. In practice this can be done by determining the so-called apparent speed c_{app} of the sound rays. Consider two

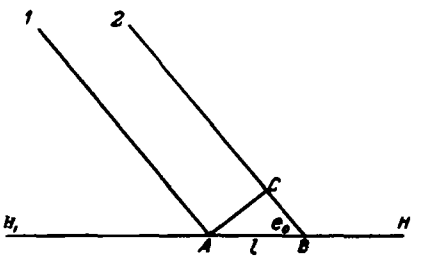


FIGURE 212

parallel rays striking the earth's surface at two points separated by the distance l (Figure 212). The apparent speed c_{app} is obtained by dividing the distance l by the time t equal to the difference in the arrival times of the rays 1 and 2 at the points A and B, i.e., $c_{app} = \frac{l}{t}$.

The quantity c_{app} can be determined from observations. Then, as one can see from the triangle ABC,

$$l = AB = tc_{app} = \frac{CB}{\cos i_0} = \frac{c_0 t}{\cos i_0}.$$

Recalling (12), we find that

$$\bar{c}_z = c_{app} = \frac{l}{t}.$$

By carrying out a series of observations at different distances from the source of sound one can construct the functional dependence between the distance and the travel time of the ray from the point of explosion, the so-called travel-time curve. Then with the aid of this curve one can represent

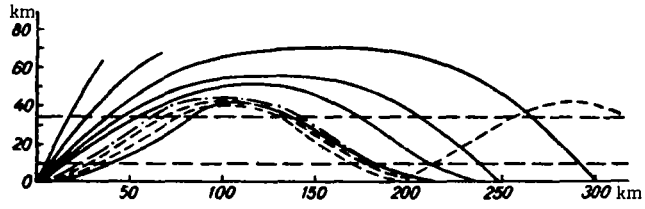


FIGURE 213. Path of sound ray forming the outer audibility zone

the course of sound rays in the atmosphere returning to the earth's surface. In reality the influence of wind on the speed of sound greatly complicates the problem. Without going into further detail we note that empirical travel-time curves have been used to determine the speed of sound c_z at different heights. All such attempts lead to the conclusion that the speed of sound decreases vertically within the troposphere ($z \approx 10$ km) only, subsequently remaining constant to a height of about 30 km; at greater altitudes, to about 70 km, it increases. For illustration we present results of such calculations:

z (km)	0	10–25	30	40	50	60	70
c_z (m/sec)	335	295	300	340	370	390	410

Figure 213 gives the average path of sound rays forming the outer audibility zone as calculated for an immobile atmosphere for a certain average state. Thus on the basis of observations one can conclude that the speed of sound c_z increases with height starting from a certain altitude. To explain this fact we turn to formula (7) of Chapter 33:

$$c = \sqrt{\varkappa \frac{R^* T}{\mu}}.$$

The above shows that the increase in the speed of sound may be produced either by a rise in the temperature T or by a reduction in the molecular weight of air μ. However, it is known that the molecular weight of air does not decrease at the altitudes in question. Therefore on the basis of acoustic data one can conclude that above 25–30 km the temperature increases with height. This conclusion agrees with data obtained by other methods. Thus the acoustic method makes it possible to obtain data relating to the distribution of temperature in the atmosphere up to the heights reached by sound waves (of the order of 60 km).

APPENDIX I

Table 1

Principal geophysical quantities and physical constants

Equatorial radius of the earth $a = 6378{,}388$ km
Polar semi-axis $b = 6356{,}911$ km
Mean radius of the earth 6371,22 km
Mean length of one degree of meridional arc 111,2 km
Surface of the earth $510 \cdot 10^6$ km^2
Volume of the earth $1083 \cdot 10^9$ km^3
Mean density 5,517 g/cm^3
Mass of the solid earth $5{,}98 \cdot 10^{21}$ tons
Mean distance of the earth from the sun $149{,}5 \cdot 10^6$ km
Mean velocity of the earth in orbit 29,77 km/sec
Speed of diurnal rotation of a point on the equator 465 m/sec
Angular velocity of diurnal rotation of the earth $\omega = 7{,}29 \cdot 10^{-5}$ sec^{-1}
Stellar day: 23 hr 56 min 4.1 sec of mean solar (civil) time
Mean day: 24 hr = 1440 min = 86400 sec.
Year: 365 days 5 hr 48 min 46 sec = 365,2422 days
Surface of oceans and seas = $362 \cdot 10^6$ km^2 (71 % of whole surface of earth)
Water stored in oceans and seas $1336 \cdot 10^6$ km^3; $1{,}3 \cdot 10^{18}$ tons
polar ice $3{,}5 \cdot 10^6$ km^3; $1{,}3 \cdot 10^{15}$ m
lakes $250 \cdot 10^3$ km^3; $0{,}25 \cdot 10^{15}$ m
underground lakes $250 \cdot 10^3$ km^3; $0{,}25 \cdot 10^{15}$ m
rivers $5 \cdot 10^3$ km^3; $50 \cdot 10^{12}$ m
swamps $6 \cdot 10^3$ km^3; $6 \cdot 10^{12}$ m
snow cover 250 km^3; $250 \cdot 10^9$ m

Hydrologic balance:

Evaporation from oceans and seas	$448 \cdot 10^3$ km^3	(86,3%)
Evaporation from land	$71 \cdot 10^3$ km^3	(13,7%)
Total	$519 \cdot 10^3$ km^3	(100%)
Precipitation over oceans and seas	$412 \cdot 10^3$ km^3	(79,4%)
Precipitation over land	$107 \cdot 10^3$ km^3	(20,6%)
Total	$519 \cdot 10^3$ km^3	(100%)

Surface run-off into oceans and seas $36 \cdot 10^3$ km^3
Mass of the atmosphere $5{,}3 \cdot 10^{15}$
Thermal balance of the earth-atmosphere system; see Table 53, p. 272.
Solar constant $S_0 = 1{,}88$ cal/cm$^2 \cdot$ min
Acceleration of gravity at latitude 45° and sea level $g_{45} = 980{,}616$ *cm/sec*^{2}
Composition and characteristics of dry air: see Table 3, p.16 and Table 4, p.17.
Characteristics of humid air: see Table 6, p.22.
Units of pressure measurement: 1 mb = 3/4 mm Hg; 1000 mb = 750,06 mm Hg; 1 mm = 4/3 mb
Normal pressure (standard atmosphere) 760 mm Hg = 1,013246 dyn/cm^2 = 1013,246 mb
1 calorie = $4{,}186 \cdot 10^7$ erg
Mechanical equivalent of heat $J = 4{,}186$ joule = $4{,}186 \cdot 10^7$ erg/cal
1 erg = $2{,}38844 \cdot 10^{-8}$ cal
Thermal equivalent of work $A = 2{,}38844 \cdot 10^{-8}$ cal/erg

Universal gas constant $R^* = 8{,}31436 \cdot 10^7$ erg/mole · deg = 1,98583 cal/mole · deg.
Specific gas constant of dry air $R_{air} = 2{,}8704 \cdot 10^6$ erg/g · deg = $6{,}8557 \cdot 10^{-2}$ cal/g · deg = $2{,}8704 \cdot 10^3$ (for p in mb)
Specific gas constant of water vapor $R_{vap} = 4{,}6150 \cdot 10^6$ erg/g · deg = 1,10226 cal/g · deg = $4{,}6150 \cdot 10^3$ (for e in mb)
Volume of one mole of ideal gas (at 0°, 760 mm Hg) $22{,}4146 \cdot 10^3$ cm³
Molecular weight of dry air 28,966
water vapor 18,016
Speed of light $c = 2{,}99792 \cdot 10^{10}$ cm/sec
Density of mercury (at 0°, 760 mm Hg) 13,5951 g/cm³
Avogadro's number $N_0 = 6{,}0247 \cdot 10^{23}$/mole
Loschmidt number (0°, 760 mm Hg) $N_L = 2{,}6875 \cdot 10^{19}$ cm³
Boltzmann constant $k = 1{,}380 \cdot 10^{-16}$ erg/deg
Planck's constant $h = 6{,}6254 \cdot 10^{-27}$ erg · sec
Stefan-Boltzmann constant $\sigma = 5{,}670 \cdot 10^{-5}$ erg/cm² · deg⁴ · sec
Electron charge $e = 4{,}80294 \cdot 10^{-10}$ CGS units
Micron $\mu = 0{,}001$ mm $= 10^{-4}$ cm
Millimicron $m\mu = 0{,}001\mu = 10^{-7}$ cm
Ångström Å $= 0{,}1\, m\mu = 10^{-8}$ cm

A few mathematical quantities

$\pi = 3{,}14159$; $\pi^2 = 9{,}8696$; $4\pi = 12{,}566$; $4/3\pi = 4{,}18879$
$e = 2{,}71828$; $e^{-1} = 0{,}36788$; $\exp a = e^a$
Constant of transition from natural to common logarithms $M = \log e = 0{,}434294$ ($\log a = M \ln a$)
Constant of transition from common to natural logarithms $\frac{1}{M} = \ln 10 = 2{,}302585$ ($\ln a = \frac{1}{M} \log a$)
Radian = 57,296° = 57°17'45"
1° = 0,01745 radian

Angle φ, deg. . . .	0	10	20	30	40	50	60	70	80	90
sin φ	0,000	0,174	0,342	0,500	0,643	0,766	0,866	0,940	0,985	1,000
cos φ	1,000	0,985	0,940	0,866	0,766	0,643	0,500	0,342	0,174	0,000

APPENDIX II

Classification of clouds

	Form	Classes	Principal subclasses	Height of base	Thickness	Microstructure	Optical phenomena, transparency	Precipitation	Pec
High clouds	I. Cirrus (Ci)	1. Ci filosus (Ci fil.)	a) Clawlike Ci uncinus (Ci unc.) b) Ridged Ci vertebratus (Ci vert.) c) Tangled Ci intortus (Ci int.)	7-10 km	From hundreds of meters to several kilometers	Crystalline. Columnar crystals, often with air gas	Sun, moon, stars shine through clearly. Blue sky shows in daytime. Parts of halo seen (rarely)	Hardly ever reaches ground	Also obse a) Ci basi cirrus clo lation of not separa gap. b) Ci radi cirrus clo convergir perspecti
		2. Ci spissatus (Ci sp.)	a) Storm (post-storm) Ci incus-genitus (Ci ing.) b) Flakelike Ci floccus (Ci floc.)						
	II. Cirrocumulus (Cc)	1. Cc undulatus (Cc und.) 2. Cc cumuliformis (Cc cuf.)	a) Lenslike Cc lenticularis (Cc lent.) a) Flakelike Cc floccus (Cc floc.) b) Cloud tracks behind aircraft Cc tractus (Cc trac.)	6-8 km	0.2-0.4 km	Crystalline. Columnar crystals, hollow prisms, separate or several combined (in complexes)	Sun, moon, stars shine through clearly. Blue sky shows in daytime.	None	Regular and thin
	III. Cirrostratus (Cs)	1. Cs filosus (Cs fil.) 2. Cs nebulosus (Cs. neb.)	None	6-8 km	From 0.1 to several kilometers	Crystalline. Cubic crystals, sometimes combined in complexes. Less often thick plates	Sun, moon, sometimes bright stars show through; blue sky seen faintly. Clear halo. Lower perihelion seen when observed from above	Hardly ever reaches ground	Sometim masses o Boundary times se sky.
Middle clouds	IV. Altocumulus (Ac)	1. Ac undulatus (Ac und.)	a) Translucent Ac translucidus (Ac trans.) b) Opaque Ac opacus (Ac op.) c) Lenslike Ac lenticularis (Ac lent.) d) Inhomogeneous Ac inhomogenus (Ac lent.)	2-6 km	0.2-0.7 km	Predominantly liquid. Droplet radius most often 5-7μ with fluctuations between 3 and 24μ. Rarely crystalline. Crystals are thick plates, less often columns or column complexes	In thin altocumulus sun and moon show through in places, corona seen. Sometimes irisation is seen	None. Precipitation bands (rare)	Often dis allel row verge du Regular seen
		2. Ac cumuliformis (Ac cuf.)	a) Flakelike Ac floccus (Ac floc.) b) Towering Ac castellatus (Ac cast.) c) Formed from cumulus Ac cumulogenitus (Ac cug.) d) With precipitation bands Ac virga (Ac vir.)	Dense to 2 km					

	Form	Classes	Principal subclasses	Height of base	Thickness	Microstructure	Optical phenomena, transparency	
Middle clouds	V. Altostratus (As)	1. As nebulosus (As neb.)	a) Translucent As translucidus (As trans.) b) Opaque As opacus (As op.) c) Precipitating As praecipitans (As pr.)	3-5 km, rarely lower	From 1 to 2 km	Mixed. Clouds consist of ice crystals and water droplets, with admixture of raindrops (at positive temperatures) or snowflakes (at negative temperatures) in lower portion of layer	Sun and moon seen as through frosted glass. Corona sometimes seen in thin translucent As. Clouds often have grayish-blue tinge.	P I A I u g
		2. As undulatus (As. und.)	Same subclasses					
	VI. Stratocumulus (Sc)	1. Sc undulatus (Sc und.)	a) Translucent Sc translucidus (Sc trans.) b) Opaque Sc opacus (Sc op.) c) Lenslike Sc lenticularis (Sc. lent.)	0.6-1.5 km, rarely lower	0.2-0.8 km	Liquid. Droplet radius mostly 5-7μ with fluctuations from 1 to 60μ. Rare ice crystals (usually plates) or snowflakes below zero.	Sun and moon can be seen only through thin cloud edges. Corona rare	G sl p
		2. Sc cumuliformis (Sc cuf.)	a) Towering Sc castellatus (Sc cast.) b) Daytime dissolving					

APPENDIX II

Classification of clouds

Thickness	Microstructure	Optical phenomena, transparency	Precipitation	Peculiarities of disposition	Distinguishing characteristics
om hundreds meters to veral kilo-eters	Crystalline. Columnar crystals, often with air gas	Sun, moon, stars shine through clearly. Blue sky shows in daytime. Parts of halo seen (rarely)	Hardly ever reaches ground	Also observed: a) Ci basis (Ci bas.)–base of cirrus clouds, dense accumulation of cirrus near horizon not separated from latter by gap. b) Ci radiatus (Ci rad.)–radial cirrus clouds in parallel bands converging at horizon due to perspective	Isolated white fibrous clouds, usually very thin and transparent but sometimes with thicker or flocculent formations. Easily distinguishable from middle clouds by their thinness, transparency and marked fibrous structure. Distinguishable from Cc by absence of undulating or tuftlike structure, from Cs by the fact that they do not form a dense veil. At twilights may been seen lighted by sun for long time. When darkness comes appear dense and dark.
2–0.4 km	Crystalline. Columnar crystals, hollow prisms, separate or several combined (in complexes)	Sun, moon, stars shine through clearly. Blue sky shows in daytime.	None	Regular rolls or light ripple and thin crests	Thin white clouds consisting of very small waves, flakes or ripples (without gray shading). Partly have fibrous structure or directly grade into Ci or Cs cover. Rarely seen separate from these clouds. Their transparency and thinness, association with typical cirrus clouds and small size of elements (waves) distinguish them from altocumulus. Also characteristic of Cs is rapid formation and rapid dissolution.
om 0.1 to veral kimeters	Crystalline. Cubic crystals, sometimes combined in complexes. Less often thick plates	Sun, moon, sometimes bright stars show through; blue sky seen faintly. Clear halo. Lower perihelion seen when observed from above	Hardly ever reaches ground	Sometimes observed in large masses obscuring entire sky. Boundary of cloud layer sometimes seen sharply against sky.	Cirrostratus clouds produce white or bluish fine uniform veil, sometimes of slightly fibrous structure. Distinguishable from cirrus in that the Cs veil is more uniform and continuous. Distinguishable from As by their lower density and presence of halo. By day under Cs objects on ground give shade.
2–0.7 km	Predominantly liquid. Droplet radius most often 5–7μ with fluctuations between 3 and 24μ. Rarely crystalline. Crystals are thick plates, less often columns or column complexes	In thin altocumulus sun and moon show through in places, corona seen. Sometimes irisation is seen	None. Precipitation bands (rare)	Often disposed in regular parallel rows seeming to converge due to perspective. Regular structure sometimes seen	White, sometimes grayish or bluish, clouds shaped like waves (ridges) consisting of individual plates or flakes. Usually these plates or flakes are separated by patches of blue sky but sometimes they fuse in a nearly solid mantle. Distinguishable from cirrocumulus by their greater density and larger apparent size of individual elements, and also by the fact that Ac is not directly connected to cirrus layers. Distinguishable from stratocumulus by their greater altitude and greater transparency.

Thickness	Microstructure	Optical phenomena, transparency	Precipitation	Peculiarities of disposition	Distinguishing characteristics
rom 1 to km	Mixed. Clouds consist of ice crystals and water droplets, with admixture of raindrops (at positive temperatures) or snowflakes (at negative temperatures) in lower portion of layer	Sun and moon seen as through frosted glass. Corona sometimes seen in thin translucent As. Clouds often have grayish-blue tinge.	Precipitation present. In winter even thin As trans. give snow. In summer rainfall usually does not reach ground.	May occur in frontal cloud system (behind the cirrostratus; give way to nimbostratus). Sometimes thin As occur separately.	Gray or bluish uniform veil of slightly fibrous structure. As a rule this veil gradually covers the entire sky. Sometimes faint waves may be seen on lower surface of the As veil. Distinguishable from cirrostratus by their greater density and the presence of precipitation. Latter usually does not reach ground but may be seen as precipitation bands against the background of As cover. Easily distinguishable from altocumulus because they form a continuous layer with no regular undulating structure. Distinguishable from Ns by their greater altitutde and lower density.
.2–0.8 km	Liquid. Droplet radius mostly 5–7μ with fluctuations from 1 to 60μ. Rare ice crystals (usually plates) or snowflakes below zero.	Sun and moon can be seen only through thin cloud edges. Corona rare	Generally none. Light short precipitation possible from Sc op.	Often disposed in parallel regular rows or waves (radial stratocumulus – Sc rad.)	Gray clouds composed of great ridges (waves), plates or flakes separated by gaps or fusing in a solid gray wavy cover. Sc clouds have much in common with Ac and are distinguishable from latter by their lower altitude, large size of the individual blocks and plates, and by their greater density. Distinguishable from stratus and nimbostratus by more pronounced lower boundary and clear undulating structure, and from Ns also by the absence of prolonged precipitation.

	Form	Classes	Principal subclasses	Height of base	Thickness	Microstructure	Optical phenomena, transparency	Precipitation
Middle clouds	V. Altostratus (As)	1. As nebulosus (As neb.) 2. As undulatus (As. und.)	a) Translucent As translucidus (As trans.) b) Opaque As opacus (As op.) c) Precipitating As praecipitans (As pr.) Same subclasses	3-5 km, rarely lower	From 1 to 2 km	Mixed. Clouds consist of ice crystals and water droplets, with admixture of raindrops (at positive temperatures) or snowflakes (at negative temperatures) in lower portion of layer	Sun and moon seen as through frosted glass. Corona sometimes seen in thin translucent As. Clouds often have grayish-blue tinge.	Precipitation presen In winter even thin As trans. give snow In summer rainfall usually does not rea ground.
	VI. Stratocumulus (Sc)	1. Sc undulatus (Sc und.) 2. Sc cumuliformis (Sc cuf.)	a) Translucent Sc translucidus (Sc trans.) b) Opaque Sc opacus (Sc op.) c) Lenslike Sc lenticularis (Sc. lent.) a) Towering Sc castellatus (Sc cast.) b) Daytime dissolving Sc diurnalis (Sc diur.) c) Evening dissolving Sc vesperalis (Sc vesp.) d) Udderlike Sc mammatus (Sc mam.)	0.6-1.5 km, rarely lower	0.2-0.8 km	Liquid. Droplet radius mostly 5-7μ with fluctuations from 1 to 60μ. Rare ice crystals (usually plates) or snowflakes below zero.	Sun and moon can be seen only through thin cloud edges. Corona rare	Generally none. Lig short precipitation possible from Sc op
Low clouds	VII. Stratus (St)	1. St nebulosus (St neb.) 2. St undulatus (St und.) 3. St fractus (St. fr.)	a) Broken rain cloud Fractonimbus (Frnb)	0.1-0.7 km	0.2-0.8 km	Predominantly liquid Droplet radius mostly 2-5μ with fluctuations from 1 to 29μ. Rare ice crystals (usually plates) or snowflakes seen at negative temperatures.	Sun and moon do not usually shine through. Bright corona sometimes seen in very thin layers.	Generally none. Occasional drizzle i summer and scatter snow or isolated fin snow grains at nega tive temperatures
	VIII. Nimbostratus (Ns)	None	None	0.1-1 km and sometimes lower	To several kilometers	Mixed. Crystals: mainly columnar at cloud top, laminar below. Droplets: radius mainly 7-8μ with fluctuations from 2 to 72μ. Clouds also contain snowflakes and raindrops	Sun and moon not seen, no particular optical phenomena	Steady rain or snow sometimes with inte ruptions.
Clouds of vertical development	IX. Cumulus (Cu)	1. Cu humilis (Cu hum.) 2. Cu mediocris (Cu. med.) 3. Cu congestus (Cu cong.)	a) Broken cumulus Cu fractus (Cu fr.) a) With cover Cu pileus (Cu pil.)	0.8-1.5 km, in certain cases considerably higher, rarely lower	From hundreds of meters to several kilometers	Liquid. Droplets largest at center and top (predominant radius 11μ), smallest near base (predominant radius 6μ)	Sun does not shine through central part of cumulus clouds, shines only through edges	Usually none. Rare isolated raindr In subtropical zones rain sometimes fall from heavy Cu
	X. Cumulonimbus (Cb)	1. Cb calvus (Cb calv.) 2. Cb capillatus (Cb. cap.)	a) With storm rolls Cb calvus arcus (Cb calv. arc.) b) With storm rolls Cb capillatus arcus (Cb. cap. arc.) c) Anvil Cb incus (Cb inc.) d) Thin Cb humilis (Cb hum.)	0.4-1.0 km sometimes lower	To several kilometers, sometimes reach tropopause	Liquid in lower region, crystalline in upper region Crystals mainly plates at high temperatures (over −15°), mainly columns at low temperatures. Snowflakes and fine graupel may also be seen.	Sun does not shine through, shaded sections dark, sections lighted by sun bright white	Torrential rain, hai graupel, etc. Preci tation bands often seen under clouds.

Microstructure	Optical phenomena, transparency	Precipitation	Peculiarities of disposition	Distinguishing characteristics
Mixed. Clouds consist of ice crystals and water droplets, with admixture of raindrops (at positive temperatures) or snowflakes (at negative temperatures) in lower portion of layer	Sun and moon seen as through frosted glass. Corona sometimes seen in thin translucent As. Clouds often have grayish-blue tinge.	Precipitation present. In winter even thin As trans. give snow. In summer rainfall usually does not reach ground.	May occur in frontal cloud system (behind the cirrostratus; give way to nimbostratus). Sometimes thin As occur separately.	Gray or bluish uniform veil of slightly fibrous structure. As a rule this veil gradually covers the entire sky. Sometimes faint waves may be seen on lower surface of the As veil. Distinguishable from cirrostratus by their greater density and the presence of precipitation. Latter usually does not reach ground but may be seen as precipitation bands against the background of As cover. Easily distinguishable from altocumulus because they form a continuous layer with no regular undulating structure. Distinguishable from Ns by their greater altitutde and lower density.
Liquid. Droplet radius mostly 5-7μ with fluctuations from 1 to 60μ. Rare ice crystals (usually plates) or snowflakes below zero.	Sun and moon can be seen only through thin cloud edges. Corona rare	Generally none. Light short precipitation possible from Sc op.	Often disposed in parallel regular rows or waves (radial stratocumulus – Sc rad.)	Gray clouds composed of great ridges (waves), plates or flakes separated by gaps or fusing in a solid gray wavy cover. Sc clouds have much in common with Ac and are distinguishable from latter by their lower altitude, large size of the individual blocks and plates, and by their greater density. Distinguishable from stratus and nimbostratus by more pronounced lower boundary and clear undulating structure, and from Ns also by the absence of prolonged precipitation. Distinguishable from cumulus by length of individual rolls and absence of considerable vertical development (except for towering Sc catellatus)
Predominantly liquid Droplet radius mostly 2-5μ with fluctuations from 1 to 29μ. Rare ice crystals (usually plates) or snowflakes seen at negative temperatures.	Sun and moon do not usually shine through. Bright corona sometimes seen in very thin layers.	Generally none. Occasional drizzle in summer and scattered snow or isolated fine snow grains at negative temperatures	None	Form solid gray layer, reminiscent of fog above ground. Lower surface of St layer is often torn, tattered. Usually the St blankets entire sky with gray veil but sometimes seen as torn cloud masses. Distinguishable from Sc by irregular structure, faint undulations and lower altitude. Distinguishable from Ns by lighter shade of gray, less pronounced fibrous structure and absence of steady rain. Further, Ns clouds are usually associated with As (with coming of front). Subclass Frnb forms only under precipitating layer (As, Ns, Sc, op.), but is not itself precipitating. Frnb layer is only pierced by precipitation of overlying clouds.
Mixed. Crystals: mainly columnar at cloud top, laminar below. Droplets: radius mainly 7-8μ with fluctuations from 2 to 72μ. Clouds also contain snowflakes and raindrops	Sun and moon not seen, no particular optical phenomena	Steady rain or snow, sometimes with interruptions.	Foul-weather fractonimbus often forms under nimbostratus clouds	Form dark gray cloud layer. When precipitating seems uniform; between periods of precipitation nonuniformity and even some undulations of cloud layer may be seen. Base of cloud layer always torn. Frnb often formed under Ns layer. Distinguishable from St clouds by darker and bluer color, nonuniformity of structure and presence of steady precipitation
Liquid. Droplets largest at center and top (predominant radius 11μ), smallest near base (predominant radius 6μ)	Sun does not shine through central part of cumulus clouds, shines only through edges	Usually none. Rare isolated raindrops. In subtropical zones rain sometimes falls from heavy Cu	Sometimes disposed in rows. Rise to considerable altitudes. If present inversion delays upward development of Cu	Dense vertically developed clouds with white dome-shaped or heaped tops and flat grayish or bluish bases. May appear against sky as isolated clouds or as considerable accumulation covering nearly entire sky. Unlike Sc, Cu do not form continuous long rolls or solid layers. Difference between Cu cong. and Ns is that tops of the latter become glaciated, assuming a cirriform structure. Ns clouds are usually precipitating or with precipitation bands hanging down.
Liquid in lower region, crystalline in upper region Crystals mainly plates at high temperatures (over −15°), mainly columns at low temperatures. Snowflakes and fine graupel may also be seen.	Sun does not shine through, shaded sections dark, sections lighted by sun bright white	Torrential rain, hail, graupel, etc. Precipitation bands often seen under clouds.	Disposed either in separate large masses or (along cold fronts) in a nearly continuous high wall	White dense clouds with dark sometimes bluish bases rising in enormous mountainlike masses. Their tops are mostly fibrous. Often seen as a few scattered clouds, but accumulation of Cb clouds or even a Cb roll may also be seen. Distinguishable from Cu cong. by spreading of top which loses rounded shape and assumes fibrous structure (at first only in places), later changing into cirriform structure which spreads over entire top part of cloud; latter often takes on anvil shape

BIBLIOGRAPHY

1. ZVEREV, A.S., KIRYUKHIN, B.V., KONDRAT'EV, K.Ya., SELEZNEVA, E.S., TVERSKOI, P.N. and M.I. YUDIN. Kurs meteorologii (fizika atmosfery) (A Course in Meteorology — Physics of the Atmosphere), edited by Prof. P.N. Tverskoi.— Gidrometeoizdat, Leningrad. 1951.

2. MATVEEV, L.T. Obshchii kurs meteorologii (osnovy fiziki atmosfery) (General Course in Meteorology — Fundamentals of Atmospheric Physics), Part I.— LKVVIA im. A.F. Mozhaiskogo, Leningrad. 1958.

3. KHRGIAN, A.Kh. Fizika atmosfery (Physics of the Atmosphere).— GITTL, Moskva. 1958..

4. OBOLENSKII, V.N. Kurs meteorologii (A Course in Meteorology).— Gidrometeoiz dat, Leningrad. 1944.

5. MITRA, S.K. The Upper Atmosphere.— The Asiatic Society, Calcutta. 1952.

6. HALTINER, J. and D.F. MARTYN. Dynamic and Physical Meteorology. [Russian translation, 1960.]

7. GANDIN, L.S., LAIKHTMAN, D.L., MATVEEV, L.T. and M.I. YUDIN. Osnovy dinamicheskoi meteorologii (Fundamentals of Dynamic Meteorology), edited by D.L. Laikhtman and M.I Yudin.— Gidrometeoizdat, Leningrad. 1955.

8. ZVEREV, A.S. Sinopticheskaya meteorologiya (Synoptic Meteorology).— Gidrometeoizdat, Leningrad. 1957.

9. ALISOV, B.P., DROZDOV, O.A. and E.S. RUBINSHTEIN. Kurs klimatologii (A Course in Climatology), edited by E.S. Rubinshtein, Parts I and II.— Gidrometeoizdat, Leningrad. 1952.

10. KEDROLIVANSKII, V.N. and M.S. STERNZAT. Meteorologicheskie pribory (Meteorological Instruments).— Gidrometeoizdat, Leningrad. 1953.

11. KALINOVSKII, A.B. and N.Z. PINUS. Aerologiya (Aerology), Part I. Metody aerologicheskikh izmerenii (Methods of Aerological Measurement).— Gidrometeoizdat, Leningrad. 1961.

12. KHRGIAN, A.Kh. Ocherki razvitiya meteorologii (On the Development of Meteorology).— Gidrometeoizdat, Leningrad. 1959.